1   **Homogenization of sub-genome secretome gene expression patterns in the**

2   **allodiploid fungus *Verticillium longisporum***

3

4   Jasper R.L. Depotter[1,2], Fabian van Beveren[1¶], Luis Rodriguez-Moreno[1¶], Grardy C.M. van

5   den Berg[1], Thomas A. Wood[2&], Bart P.H.J. Thomma[1&*], Michael F. Seidl[1&*]

6

7   [1]Laboratory of Phytopathology, Wageningen University & Research, Wageningen, The

8   Netherlands

9   [2]Department of Crops and Agronomy, National Institute of Agricultural Botany, Cambridge,

10   United Kingdom

11

12   *Corresponding authors

13   E-mail: bart.thomma@wur.nl (BPHJT)

14   E-mail: michael.seidl@wur.nl (MFS)

15

16   ¶These authors contributed equally to this work

17   &These authors also contributed equally to this work

Abstract

Allopolyploidization, genome duplication through interspecific hybridization, is an important evolutionary mechanism that can enable organisms to adapt to environmental changes or stresses. The increased adaptive potential of allopolyploids can be particularly relevant for plant pathogens in their ongoing quest for host immune response evasion. To this end, plant pathogens secrete a plethora of molecules that enable host colonization. Allodiploidization has resulted in the new plant pathogen *Verticillium longisporum* that infects different hosts than haploid *Verticillium* species. To reveal the impact of allodiploidization on plant pathogen evolution, we studied the genome and transcriptome dynamics of *V. longisporum* using next-generation sequencing. *V. longisporum* genome evolution is characterized by extensive chromosomal rearrangements, between as well as within parental chromosome sets, leading to a mosaic genome structure. In comparison to haploid *Verticillium* species, *V. longisporum* genes display stronger signs of positive selection. The expression patterns of the two sub-genomes show remarkable resemblance, suggesting that the parental gene expression patterns homogenized upon hybridization. Moreover, whereas *V. longisporum* genes encoding secreted proteins frequently display differential expression between the parental sub-genomes in culture medium, expression patterns homogenize upon plant colonization. Collectively, our results illustrate of the adaptive potential of allodiploidy mediated by the interplay of two sub-genomes.

2

38 **Author summary**

39 Hybridization followed by whole-genome duplication, so-called allopolyploidization,

40 provides genomic flexibility that is beneficial for survival under stressful conditions or

41 invasiveness into new habitats. Allopolyploidization has mainly been studied in plants, but

42 also occurs in other organisms, including fungi. *Verticillium longisporum*, an emerging fungal

43 pathogen on brassicaceous plants, arose by allodiploidization between two *Verticillium* spp.

44 We used comparative genomics to reveal the plastic nature of the *V. longisporum* genomes,

45 showing that parental chromosome sets recombined extensively, resulting in a mosaic genome

46 pattern. Furthermore, we show that non-synonymous substitutions frequently occurred in *V.

47 longisporum*. Moreover, we reveal that expression patterns of genes encoding secreted

48 proteins homogenized between the *V. longisporum* sub-genomes upon plant colonization. In

49 conclusion, our results illustrate the large adaptive potential upon genome hybridization for

50 fungi mediated by genomic plasticity and interaction between sub-genomes.

Introduction

53 Cycles of polyploidization are hallmarks of eukaryotic genome evolution, where an initial

54 increase of ploidy is commonly followed by reversions to original ploidy states [1]. For

55 instance, all angiosperm plants share two rounds of ancient polyploidy [2]. The prevalence of

56 polyploidy events in eukaryotic evolution is likely due to the high evolutionary potential of

57 polyploids, as additional chromosome sets give leeway to functional diversification [3].

58 Consequently, polyploidy events are often followed by adaptive radiation and are dated near

59 the base of species-rich clades in phylogenies [4,5]. In addition, polyploidy has been

60 associated with increased invasiveness [6] and resistance to environmental stresses [7]. For

61 instance, numerous plant species that survived the Cretaceous-Palaeogene mass extinction 66

62 million years ago underwent a polyploidization event which is thought to have contributed to

63 their increased survival rates [8,9]. Polyploids may originate from the same species, i.e.

64 autopolyploidization, or from different species as a result of interspecific hybridization, i.e.

65 allopolyploidization. In general, allopolyploids are believed to have a higher adaptive

66 potential than autopolyploids due to the combination of novel or diverged genes from distinct

67 parental species [3].

68   The impact of allopolyploidization has mainly been investigated in plants, as

69 approximately a tenth of all plant species consists of allopolyploids [10]. In contrast,

70 allopolyploidization in fungi is far less intensively investigated [11]. Nonetheless,

71 allopolyploidization impacted the evolution of numerous fungal species, including the

72 economically important baker's yeast *Saccharomyces cerevisiae* [12]. The increased adaptive

73 potential enabled allopolyploid fungi to develop desirable traits that can be exploited in

74 industrial bioprocessing [13]. For instance, at least two recent hybridization events between *S.*

75 *cerevisiae* and its close relative *Saccharomyces eubayanus* gave rise to *Saccharomyces*

76 *pastorianus*, a species with high cold tolerance and good maltose/maltotriose utilization

4

77   capabilities, which is exploited in the production of lager beer that requires barley to be

78   malted at low temperatures [14].

79   Allopolyploid genomes experience a so-called "genome shock" upon hybridization,

80   inciting major genomic reorganizations that can manifest by genome rearrangements,

81   extensive gene loss, transposon activation, and alterations in gene expression [15]. These

82   early stage alterations are primordial for hybrid survival, as divergent evolution is principally

83   associated with incompatibilities between the parental genomes [16]. Additionally, these

84   initial re-organizations and further alterations in the aftermath of hybridization also provide a

85   source for environmental adaptation [3]. Frequently, heterozygosity is lost for many regions

86   in the allopolyploid genome [17]. This can be a result of the direct loss of a homolog of

87   different parental origin (i.e. a homeolog) through deletion or gene conversion whereby one

88   of the copies substitutes its homeologous counterpart [18]. Gene conversion and the

89   homogenization of complete chromosomes played a pivotal role in the evolution of the

90   osmotolerant yeast species *Pichia sorbitophila* [19]. In total, two of its seven chromosome

91   pairs consist of partly heterozygous, partly homozygous sections, whereas two chromosome

92   pairs are completely homozygous. Gene conversion may eventually result in chromosomes

93   consisting of sections of both parental origins as "mosaic genomes" [20]. However, mosaic

94   genomes can also arise through recombination between chromosomes of the different parents,

95   such as in the hybrid yeast *Zygosaccharomyces parabailii* [21].

96   The redundancy of having one or more additional homeolog copies for most genes

97   facilitates functional diversification in allopolyploids [22]. Consequently, accelerated gene

98   evolution is generally observed upon allopolyploidization [22,23]. Nevertheless, recent

99   allopolyploidization events in the fungal genus *Trichosporon* resulted in a general

100   deceleration of gene evolution [24]. Thus, allopolyploidization is not always followed by

101 accelerated gene evolution. Arguably, the environmental exposure upon allopolyploidization

102 plays an pivotal role in the eventual speed and grade of gene diversification [25].

103   Allopolyploidization typically entails gene expression pattern alterations when

104 parental genomes that evolved distinct transcriptional regulation merge [26]. Moreover,

105 crosstalk between parental sub-genomes can lead to further gene expression alterations,

106 leading to the emergence of novel expression patterns [27–29]. Nevertheless, in general,

107 expression patterns are often conservatively inherited upon hybridization as the majority of

108 allopolyploid genes are expressed similarly to their parental orthologs [30]. For instance,

109 more than half of the genes in an allopolyploid strain of the fungal grass endophyte *Epichloë*

110 retained their parental gene expression level [31].

111   Plant pathogens are often thought to evolve while being engaged in arms races with

112 their hosts with pathogens evolving to evade host immunity while plant hosts attempt to

113 intercept pathogen ingress [32,33]. Due to the increased adaptation potential,

114 allopolyploidization has been proposed as a potent driver in pathogen evolution [34].

115 Allopolyploids often have different pathogenic traits than their parental lineages, such as

116 higher virulence [35,36] and altered host ranges [37,38]. Within the fungal genus *Verticillium*,

117 allodiploidization resulted in the emergence of the species *Verticillium longisporum* [37,39].

118 *V. longisporum* infects brassicaceous plants, whereas other *Verticillium* spp. generally do not

119 typically colonize plants of this family, with the exception of *Arabidopsis thaliana* [37,40–

120 42]. Similar to haploid *Verticillium* spp., *V. longisporum* is thought to have predominant

121 asexual reproduction as a sexual cycle has never been described and populations are not

122 outcrossing [39,43]. *V. longisporum* is sub-divided into three lineages, each representing a

123 separate hybridization event [37]. The economically most important lineage A1/D1 originates

124 from hybridization between *Verticillium* species A1 and D1 that have hitherto not been found

125 in their haploid states. *V. longisporum* lineage A1/D1 is the main causal agent of Verticillium

126   stem striping on oilseed rape and a worldwide emerging pathogen [44,45]. Lineage A1/D1

127   can be further divided into two genetically distinct populations, which have been named

128   'A1/D1 West' and 'A1/D1 East' after their relative geographic occurrence in Europe [39].

129   Nevertheless, both populations originate from the same hybridization event [39].

130   Conceivably, upon or subsequent to the hybridization event, *V. longisporum* encountered

131   extensive genetic and transcriptomic alterations that might have facilitated a shift towards

132   brassicaceous hosts. Here, we studied the impact of allodiploidization on the evolution of *V.*

133   *longisporum* by investigating genome, gene and transcriptomic plasticity.

134

Results

V. longisporum displays a mosaic genome structure

137   The genomes of two *V. longisporum* strains were analysed to investigate the impact of

138   hybridization on the genome structure. Previously, *V. longisporum* strains VLB2 and VL20,

139   belonging to 'A1/D1 West' and 'A1/D1 East', were sequenced with the PacBio RSII platform

140   and assembled into genomes of 72.9 and 72.3 Mb in size, respectively [39]. These genome

141   sizes exceed double the amount of the telomere-to-telomere sequenced *V. dahliae* strains JR2

142   (36.2 Mb) and VdLs17 (36.0 Mb) [46]. We used RepeatModeler (V1.0.8; Smit et al. 2015) in

143   combination with RepeatMasker to determine that 14.28 and 13.90% of the *V. longisporum*

144   strain VLB2 and VL20 genomes are composed of repeats, respectively (S1 Table).

145   Intriguingly, this is more than double the repeat content as in *V. dahliae* strain JR2, for which

146   6.49% of the genome was annotated as repeat using the same methodology. The *V.*

147   *longisporum* genomes were also screened for the fungal telomere-specific repeats

148   (TAACCC/GGGTTA) to estimate the number of chromosomes. In total, 29 and 30 telomeric

149   regions were found in the VLB2 and VL20 genomes, respectively, that were consistently

150   situated at the end of sequence contigs, suggesting that *V. longisporum* contains at least 15

7

151   chromosomes (S1 Table). Six out of 45 and 4 out of 44 sequence contigs in strains VLB2 and

152   VL20, respectively, were flanked on both ends by telomeric repeats and therefore likely

153   represent complete chromosomes (S1 Table). For comparison, *V. dahliae* strains have 8

154   chromosomes [46].

155   In allodiploid organisms, parental origin determination is elementary to investigate

156   genome evolution in the aftermath of hybridization. As species D1 is phylogenetically closer

157   related, and consequently has a higher sequence identity, to *V. dahliae* than species A1, *V.*

158   *longisporum* genomic regions were previously provisionally assigned to either species D1 or

159   A1 [39]. Here, we determined the parental origin of *V. longisporum* genomic regions more

160   precisely. The difference in phylogenetic distance of species A1 and D1 to *V. dahliae* caused

161   that *V. longisporum* genome alignments to *V. dahliae* displayed a bimodal distribution with

162   one peak at 93.1% and another peak at 98.4% sequence identity that represent the two parents

163   with a dip at 96.0% (S1 Fig). In order to separate the two sub-genomes, regions with an

164   average sequence identity to *V. dahliae* of <96% were assigned to species A1, whereas

165   regions with an identity of ≥96% were assigned to species D1 (Fig 1). In this manner, 36.2

166   Mb of *V. longisporum* strain VLB2 was assigned to species A1 and 35.7 Mb to species D1.

167   For *V. longisporum* strain VL20, 36.3 Mb was assigned to species A1 and 35.2 Mb to species

168   D1. Only 1.0 and 0.8 Mb of strains VLB2 and VL20, respectively, could not be aligned to *V.*

169   *dahliae* and thus remained unassigned.

170   To trace the chromosome sets of the original parents of the hybrid, the parental origin

171   of individual contigs was determined. In total, 8 of the 10 largest contigs of *V. longisporum*

172   strain VLB2 as well as strain VL20 consist of regions originating from both species A1 and

173   species D1 (Fig 1). Thus, parental chromosome sets cannot be separated from one another as

174   *V. longisporum* apparently evolved a mosaic genome structure in the aftermath of

175   hybridization.

176

Genomic rearrangements are responsible for the mosaic genome

Typically, a mosaic structure of a hybrid genome can originate from gene conversion or from chromosomal rearrangements between DNA strands of different parental origin [17]. To analyse the extent of gene conversion, genes were predicted for the *V. longisporum* strains VLB2 and VL20. To aid gene annotation with the BRAKER1 1.9 pipeline [47], ~2 Gb of filtered RNA-seq reads were generated from fungal cultures in liquid medium. In total, 19,123 and 18,784 genes were predicted for *V. longisporum* strains VLB2 and VL20 respectively, which is ~90% higher than the amount of genes that were predicted for *V. dahliae* strain JR2 using the same approach (9,909 genes) (S1 Table). As to be expected, the divergence of species A1 and D1 could also been observed at the gene level based on sequence identity and GC-content (Fig 1; S1 Fig). In total, 9,531 and 9,402 genes were assigned to the species A1 sub-genome of the strains VLB2 and VL20, respectively, whereas the number of genes in the species D1 sub-genomes was 9,468 and 9,243 for these strains, respectively (S1 Table). Thus, the amount of genes is similar in the two sub-genomes for both *V. longisporum* strains and similar to the gene number identified in *V. dahliae* strain JR2. Over 80% of the *V. longisporum* genes are present in two copies whereas almost all genes (97-98%) are present in one copy within each of the *V. longisporum* sub-genomes (S2 Fig). Moreover, of the 7,620 genes that are present in two copies in VLB2 and VL20, only 5 genes were found to be highly similar (<1% nucleotide sequence diversity) in VLB2, whereas the corresponding gene pair in VL20 was more diverse (>1%) (Fig 2A). In *V. longisporum* strain VL20, no highly similar copies were found that are more divergent in VLB2. Collectively, these findings indicate that the two copies of most genes present in the *V. longisporum* are homeologs and that gene conversion only played a minor role during evolution of the mosaic genome.

200   Considering that gene conversion played a minor role during genome evolution, the

201   mosaic genome structure of *V. longisporum* likely originated from rearrangements between

202   homeologous chromosomes. To identify the location of genomic rearrangements, the genome

203   of *V. longisporum* strain VLB2 was aligned to that of strain VL20 (Fig 2B). Extensive

204   chromosomal rearrangements occurred between the two *V. longisporum* strains, as we

205   observed 87 putative syntenic breaks. In order to confirm these breaks, long sequencing reads

206   of VLB2 were aligned to the VL20 genome assembly to assess if synteny breaks were

207   supported by read mapping (S3 Fig). In total, 60 synteny breaks could be confirmed by read

208   mapping. As genomic rearrangements are often associated with repeat-rich genome regions

209   [48], the synteny break points were tested for their association to these regions. In total, 34 of

210   the 60 (57%) confirmed synteny break points were flanked by repeats, which is significantly

211   more than expected from random sampling (mean = 18.5%, σ = 0.05%) (S4 Fig). In

212   conclusion, it appears that chromosomal rearrangement, rather than gene conversion, is the

213   main driver underlying the mosaic structure of the *V. longisporum* genome.

214

V. longisporum lost heterozygosity through deletions

216   In each of the *V. longisporum* isolates, 17% of the genes occur only in a single copy.

217   Although gene conversion played a minor role in the aftermath of hybridization, loss of

218   heterozygosity may occur through gene loss or, alternatively, single-copy genes may originate

219   from parent-specific contributions. However, as 12% of the singly copy genes in strain VLB2

220   are present in two copies in strain VL20, and 16% of the single copy genes in VL20 are

221   present in two copies in VLB2, gene deletion seems to be an on-going process in *V.*

222   *longisporum* evolution since both strains are derived from the same hybridization event [39].

223   Of the genes that were lost in the VLB2 divergence from VL20, 52% resided in the species

224   A1 sub-genome, whereas 47% in the D1 sub-genome (1% remained unassigned). For VL20,

225   49% and 50% of these lost genes resided in the A1 and D1 sub-genome, respectively. Thus,

226   gene loss occurs evenly across the two sub-genomes. We next determined the fraction of lost

227   genes that encode secreted proteins as pathogen secretomes play pivotal roles in establishing

228   symbioses with plant hosts [49]. In total, 12.9% and 8.7% of the genes that encode secreted

229   proteins were lost in the divergence of VLB2 and VL20, respectively. This is a significant

230   enrichment for strain VLB2, where genome-wide 8.9% of the genes encode secreted proteins

231   (Fisher's exact test, P = 0.009) (S1 Table). Nevertheless, this enrichment is not found for

232   strain VL20 of which 9.0% of the genes encode secreted proteins (Fisher's exact test, P =

233   0.45) (S1 Table). Nonetheless, in general, *V. longisporum* strains VLB2 and VL20 contain

234   1.91 and 1.90 times the number of genes encoding secreted proteins compared to *V. dahliae*

235   JR2 with similar contributions of the species A1 and D1 sub-genomes (S1 Table). This

236   indicates that, although gene loss occurs, it hitherto impacted the secretome of *V. longisporum*

237   only to a limited extent.

238

239   **Global acceleration of gene evolution upon allodiploidization**

240   To investigate the evolution of genes subsequent to the allodiploidization event, we

241   determined their rates of non-synonymous (*Ka*) and synonymous (*Ks*) substitutions.

242   Substitution rates were determined for so-called best-reciprocal orthologs, which are genes

243   that are present in a single copy in all *Verticillium* species (Fig 3A) and thus, sub-genomes A1

244   and D1 of *V. longisporum* were considered separately. In total, 5,342 and 5,369 orthologous

245   groups could be constructed using *V. longisporum* strain VLB2 and VL20, respectively.

246   Consequently, for every orthologous group, *Ka/Ks* ratios were determined for every branch of

247   the *Verticillium* phylogeny leading to an extant species and consequently compared with the

248   *Ka/Ks* ratio obtained for the *V. dahliae* branch. In general, genes in clade Flavexudans spp.

249   displayed higher *Ka/Ks* ratios than clade Flavnonexudans spp. (Fig 3; S5 Fig). *V. nubilum* and

11

250   *V. albo-atrum* genes displayed the lowest *Ka/Ks* ratios of all *Verticillium* spp., which

251   correlates with their relatively long evolutionary history without divergence of known sister

252   species (Fig 3; S5 Fig). Of all *Verticillium* spp., *V. longisporum* sub-genome D1 was the only

253   species of which genes displayed significantly higher *Ka/Ks* ratios than *V. dahliae* (Wilcoxon

254   rank-sum test, *P* = 9.43e-12, VLB2 based). Thus, genes of the *V. longisporum* sub-genome

255   D1 generally evolve faster than genes of other haploid *Verticillium* spp. In contrast, genes of

256   the other *V. longisporum* sub-genome, A1, generally displayed lower *Ka/Ks* ratios than *V.*

257   *dahliae* orthologs (Wilcoxon rank-sum test, *P* < 2.2e-16, VLB2 based). However, the absence

258   of a phylogenetically closely related A1 sister species hampers the determination of putative

259   differences in gene diversification rate before and after hybridization as *Ka/Ks* ratios can vary

260   considerably between genes of haploid *Verticillium* spp. For instance, *V. longisporum* sub-

261   genome A1 displays higher evolutionary speed, expressed by *Ka/Ks*, than *V. alfalfae*

262   (Wilcoxon rank-sum test, *P* = 3.94e-14, VLB2 based) and a similar evolutionary speed to *V.*

263   *nonalfalfae* (Wilcoxon rank-sum test, *P* = 0.09, VLB2 based), species that have the same last

264   common ancestor with A1 as *V. dahliae* (Fig 3A).

265   To find evidence for accelerated evolution in both *V. longisporum* sub-genomes, we

266   determined the number of genes under positive selection in every *Verticillium* (sub-)genome

267   using the above formed orthologous groups. Genes under positive selection were determined

268   based on a *Z*-test and varied considerably from 209 in the *V. longisporum* sub-genome A1 to

269   3 in *V. tricorpus* (S6 Fig). Intriguingly, the *V. longisporum* sub-genomes A1 and D1 have the

270   highest number of genes under positive selection (128 genes for species D1). The genomes *V.*

271   *dahliae* and *V. zaregansianum* contain a considerable number of genes under positive

272   selection; 103 and 99 respectively. To investigate whether particular functional gene

273   properties are associated with genes under positive selection, the fractions of genes that

274   encode secreted proteins were determined (S6 Fig). These fractions were not higher in the *V.*

275   *longisporum* sub-genomes than in *V. dahliae* and *V. zaregansianum*. Furthermore, no major

276   differences in Clusters of Orthologous Group (COG) functional categories could be observed

277   for genes under positive selection between sub-genome A1, sub-genome D1, *V. dahliae* and

278   *V. zaregansianum* (S7 Fig). Thus, we conclude that *V. longisporum* genes globally diverge

279   faster than genes of related haploid *Verticillium* spp.

280

281   **Expression pattern homogenization in the hybridization aftermath**

282   To investigate the impact of allodiploidization on gene expression patterns, the expression of

283   *V. longisporum* genes was compared with *V. dahliae* orthologs of isolates grown in culture

284   medium. To this end, expression of single copy *V. dahliae* genes was compared with *V.*

285   *longisporum* orthologs that are present in two copies: one in the species A1 sub-genome and

286   one in the species D1 sub-genome. In total, 7,469 and 7,411 of these expressed gene clusters

287   were found for *V. longisporum* strain VLB2 and VL20, respectively. Reads were mapped to

288   the predicted *V. longisporum* genes of which 51% and 50% mapped to species A1 homeologs

289   and 49% and 50% to the species D1 homeologs, for strains VLB2 and VL20, respectively,

290   and thus we observed no general dominance in expression for one of the sub-genomes. Over

291   60% of the *V. dahliae* genes are not differentially expressed compared to A1 and D1

292   orthologs, indicating that the majority of the genes did not evolve differential expression

293   patterns (Fig 4). In total, 29.6% and 24.2% of the genes are differently expressed compared to

294   *V. dahliae* orthologs in the species A1 and D1 sub-genomes, respectively (Fig 4, VLB2

295   based). The significantly higher fraction of differentially expressed A1 genes (Fisher's exact

296   test, $P = 4.3e-13$) corresponds to the more distant phylogenetic relationship of A1 with *V.*

297   *dahliae* than of D1. Intriguingly, however, significantly less D1 genes (18.5%, VLB2 based)

298   were differently expressed to A1 homeologs than to *V. dahliae* orthologs despite the larger

299   phylogenetic distance between *Verticillium* species A1 and D1 (Fisher's exact test, $P < 2.2e-$

300   16). In general, the expression pattern of *V. longisporum* A1 and D1 sub-genomes are more

301   similar to each other (ρ = 0.90 for VLB2) than the expression pattern of sub-genome D1 and

302   *V. dahliae* (ρ = 0.86 for VLB2) (Fig 5; S2 Table). This discrepancy in phylogenetic

303   relationship and expression pattern similarities may indicate that the expression patterns of

304   species A1 and D1 homogenized upon hybridization. Moreover, for *V. longisporum* strains

305   VLB2 and VL20, homeolog expression patterns within the same strain (ρ = 0.90 for VLB2)

306   correlated more than A1 and D1 expression patterns of different strains (ρ = 0.88 for VLB2

307   A1 and VL20 D1) (Fig 5; S2 Table). Thus, expression patterns of homeologs may have

308   synchronized in the aftermath of hybridization.

309

310   ***In planta* secretome homogenization between *V. longisporum* sub-genomes**

311   To assess potential gene expression differences upon host colonization, gene expression

312   patterns of *V. longisporum* and *V. dahliae* orthologs were also investigated *in planta*. To this

313   end, oilseed rape plants were inoculated with VLB2 and VL20, respectively. As observed

314   previously, oilseed rape plants inoculated with VLB2 developed typical *Verticillium*

315   symptoms including stunted plant growth and leaf chlorosis [50]. In contrast, oilseed rape

316   plants inoculated with VL20 did not display any disease symptoms. Accordingly, VLB2 DNA

317   could be detected in oilseed rape stems, whereas VL20 DNA remained under the detection

318   limit. In addition, *A. thaliana* plants were inoculated with *V. dahliae* strain JR2, and *V.*

319   *longisporum* strains VLB2 and VL20. However, only JR2 DNA could be detected in above-

320   ground plant material. Consequently, total RNA sequencing was performed for oilseed rape

321   plants inoculated with *V. longisporum* strain VLB2 and *A. thaliana* plants inoculated with *V.*

322   *dahliae* strain JR2. In total, ~1.5 Gb of filtered RNA-seq reads were generated from the

323   *Verticillium* inoculated plant material. Similar to *V. longisporum* grown in culture medium,

324   49% and 51% of the reads mapped to the A1 and D1 sub-genomes of *V. longisporum*. Thus,

325 also *in planta* there is no expression dominance of one of the *V. longisporum* sub-genomes.

326 Furthermore, 16.4% and 15.1% of the *V. longisporum* genes were differently expressed from

327 *V. dahliae* orthologs in sub-genome A1 and D1, respectively (Fig 6). Thus, in correspondence

328 with *V. longisporum* grown in culture medium, a larger fraction of A1 orthologs was

329 differently expressed to *V. dahliae* orthologs than D1 orthologs (Fisher's exact test, $P = 0.04$).

330   To elucidate a putative association of gene expression differences between *V.*

331 *longisporum* and *V. dahliae* with their distinct host ranges, the fraction of differently

332 expressed genes that encode secreted proteins was determined. For *Verticillium* grown in

333 culture medium, 16.5% of the differentially expressed genes between *V. longisporum* and *V.*

334 *dahliae* encode secreted proteins, whereas this is only 5.3% for genes without differential

335 expression (Fig 6A). This enrichment of genes encoding secreted proteins was also found for

336 isolate VL20 (S8 Fig). Correspondingly, *V. longisporum* genes that were differently expressed

337 from *V. dahliae* orthologs were enriched for Pfam domains associated with secretion and host

338 colonization, such as *Hce2* (PF14856), a domain found in putative effectors with homology to

339 *Cladosporium fulvum* effector Ecp2 (Stergiopoulos et al. 2012; S3 Table). Similarly, genes

340 encoding secreted proteins were also significantly enriched for differentially expressed genes

341 between *V. longisporum* and *V. dahliae in planta* (Fig 6B). However, the fraction of 9.7% was

342 significantly less than 16.5% for *Verticillium* grown in culture medium (Fig 6B). Thus,

343 despite the colonization of *V. longisporum* and *V. dahliae* on a different host species, the

344 enrichment of genes encoding secreted proteins is lower for *Verticillium* grown *in planta*

345 compared with culture medium.

346   To see how the different sub-genomes contribute to the enrichment of genes encoding

347 secreted proteins, we compared gene expression patterns of *V. longisporum* sub-genomes. For

348 *V. longisporum* grown in culture medium, 19.4% of the genes that are differentially expressed

349 between the A1 and D1 homeologs encode secreted proteins, whereas this is 5.8% for

15

350 homeologs with similar expression levels (Fig 7). Thus, similar to *V. longisporum* and *V.*

351 *dahliae* orthologs, differentially expressed homeologs are enriched for genes that encode

352 secreted proteins. Intriguingly, 7.9% of the genes with differential homeolog expression *in*

353 *planta* encode secreted proteins, which is a similar fraction as for homeologs without

354 differential expression (8.4%; Fig 7). Thus, upon plant colonization, there is no enrichment of

355 genes that encode secreted proteins for differentially expressed homeologs. This lack of

356 enrichment may be due to increased expression differences *in planta* between homeologs that

357 encode non-secreted proteins. Alternatively, expression levels of homeologs encoding

358 secreted proteins may homogenize *in planta* (Fig 7). In total, 11.2% and 9.5% of the genes

359 that are differently regulated *in planta* from culture medium encode secreted proteins in sub-

360 genome A1 and D1, respectively (Fig 7). This is a significantly larger fraction than for genes

361 without differential expression: 7.2% and 7.9% for sub-genomes A1 and D1, respectively. In

362 sub-genome A1, this enrichment of genes encoding secreted proteins was both present for *in*

363 *planta* up-regulated (11.6%, Fisher's exact test, $P < 1.74e-05$) and down-regulated (10.9%,

364 Fisher's exact test, $P < 4.23e-05$) genes. In contrast, in sub-genome D1, the enrichment was

365 present for *in planta* up-regulated genes (12.3%, Fisher's exact test, $P < 2.27e-05$), but not for

366 down-regulated genes (7.7%, Fisher's exact test, $P = 0.91$). Thus, in general, genes encoding

367 secreted proteins underwent relatively more frequently expression alterations upon plant

368 colonization compared to genes that encode for non-secreted proteins. Consequently, the lack

369 of enrichment *in planta* is caused by an increased homogenization of homeolog expression

370 patterns of genes encoding secreted proteins. This homogenization is illustrated by the

371 expression pattern of genes encoding secreted proteins with a pectate lyase Pfam domain

372 (PF03211), which is associated with host cell-wall degradation (S9 Fig) [52]. In total, 5 genes

373 (*Pect_ly_1*, *Pect_ly_2*, *Pect_ly_3*, *Pect_ly_4* and *Pect_ly_5*), with a pectate lyase domain

374 were found with two homeologous copies in *V. longisporum* and all of them were predicted to

16

375   be secreted. There was no differential expression between the homeologs *in planta* for all 5

376   genes, whereas in culture medium the homeologs are differentially expressed for *Pect_ly_1*,

377   *Pect_ly_4* and *Pect_ly_5*. Homogenization of homeolog expression was achieved through

378   down-regulation for *Pec_ly_1*, as both homeologs were not expressed *in planta*. In contrast,

379   similar levels of homeolog *in planta* expression were achieved for *Pect_ly_4* and *Pect_ly_5*

380   by the relative increase in expression of the D1 and A1 homeolog, respectively.

381

382   **DISCUSSION**

383   Hybridization is a powerful evolutionary mechanisms often leading to the emergence of new

384   plant pathogens with distinct pathogenic features from their parents [34,53]. We demonstrate

385   the genomic and transcriptomic plasticity of the allodiploid *V. longisporum* pathogen and

386   illustrate its potential for divergent evolution. Firstly, the plastic nature of the *V. longisporum*

387   genome is displayed by its mosaic structure (Fig 1). Mosaicism in *V. longisporum* is not

388   driven by homogenization that played a negligible role in the aftermath of hybridization (Fig

389   2A). Rather, *V. longisporum* mosaic genome structure is caused by extensive genomic

390   rearrangements after hybridization (Fig 2B). Genomic rearrangements are major drivers of

391   evolution and facilitate adaptation to novel or changing environments [48]. Genomic

392   rearrangements are not specific to the hybrid nature of *V. longisporum* as other *Verticillium*

393   spp. similarly encountered extensive chromosomal reshuffling [54–56]. As expected, the

394   majority of the synteny breaks between the genomes of *V. longisporum* strains VLB2 and

395   VL20 reside in repeat-rich genome regions (S4 Fig) as, due to their abundance, repetitive

396   sequences are more likely to act as a substrate for unfaithful repair of double-strand DNA

397   breaks [48]. Nonetheless, in *V. longisporum*, 43% of the synteny breaks identified are not

398   associated with repeat-rich regions. Conceivably, the presence of two genomes also provides

399   homeologous sequences with sufficient identity to mediate unfaithful repair. Secondly, *V.*

17

400 *longisporum* genes globally display accelerated evolution in comparison to orthologs of non-

401 hybrid *Verticillium* spp. This is illustrated by the increased abundance of genes under positive

402 selection and the more divergent evolution of D1 genes in comparison to its sister species *V.*

403 *dahliae* (Figs 3; S5 and S6 Figs). The increased rate of divergence is likely a result of having

404 two homeologs of most genes as this redundancy gives leeway to functional diversification

405 [57]. Previously, 29 genes in *V. dahliae* were determined to evolve under positive selection,

406 whereas in this study more than three times the amount was found (S6 Fig) [54]. Here, we

407 obtained higher numbers as a higher *P*-value cut-off was used ($P < 0.05$ and instead of $P <$

408 0.01). Moreover, we calculated positive selection based on the nucleotide substitutions along

409 the *V. dahliae* species branch, whereas de Jonge et al. (2013) determined positive selection

410 based on intraspecific substitutions, which is expected to have lower *Ka/Ks* ratios for genes

411 under positive selection than when interspecific mutations are used [58]. Finally,

412 allodiploidization resulted in transcriptomic alterations even though the majority of the *V.*

413 *longisporum* genes was not differently expressed from *V. dahliae* orthologs (Fig 4). As

414 species A1 and D1 are hitherto unfound in their haploid state, *V. dahliae* was used for

415 expression pattern comparison as it only recently diverged from species D1 and therefore

416 likely resembles D1 gene expression [37]. Unfortunately, species A1 currently lacks a known

417 sister species that could be used in the gene expression comparison (Fig 3A). Despite the

418 absence of the *V. longisporum* parents, expression patterns between the A1 and D1 sub-

419 genomes seem to have homogenized upon hybridization as they show more resemblance than

420 the expression pattern between *V. dahliae* and species D1 (Fig 5; S2 Table). *V. longisporum*

421 genes that were differently expressed to *V. dahliae* orthologs are enriched for genes encoding

422 secreted proteins (Fig 6; S8 Fig). However, that enrichment is higher for *Verticillium* grown

423 in culture medium compared with *in planta* (Fig 6). This may be a consequence of the

424 disappearance of expression differences between the *V. longisporum* homeologs as

18

425   homeologs that encode secreted proteins homogenized their expression upon host

426   colonization (Fig 7).

427   Whole-genome duplication events are usually followed by extensive gene loss, often

428   leading to reversion to the original ploidy state [59]. However, the so-called 'haploidization'

429   of *V. longisporum* has only proceeded to a limited extent, as 80% of the genes are present in

430   two copies (S2 Fig), whereas the haploid *V. dahliae* genome contains only 1% of its genes in

431   two copies. Thus, the *V. longisporum* genome displays the symptoms of a recent allodiploid,

432   with gene loss being an on-going process that by now has only progressed marginally.

433   However, the retention of both homeolog copies can also be evolutionary advantageous as the

434   presence of an additional gene copy may facilitate functional diversification of *V.*

435   *longisporum* genes in comparison to haploid *Verticillium* spp. (Fig 3; S5 and S6 Figs). Gene

436   duplication is an important mechanism for plant pathogens, including *V. dahliae*, to evade

437   host immunity [55,60,61]. The LS regions of *V. dahliae*, which are enriched for active

438   transposable elements, are derived from segmental duplications [54,55]. However, instead of

439   the specific duplication of LS regions, *V. longisporum* hybridization resulted in a whole-

440   genome duplication likely resulting in the global acceleration of gene evolution as genes

441   encoding secreted proteins or genes associated with niche colonization were not enriched in

442   genes that evolve under positive selection (S6 and S7 Figs).

443   Expression divergence of a particular gene may evolve through mutations in

444   regulatory sequences of the gene itself (cis effects), such as promoter elements, or alterations

445   in other regulatory factors (trans effects), such as chromatin regulation [27,62]. Conceivably,

446   the surprisingly higher correlation of the sub-genome D1 expression pattern with A1 than

447   with *V. dahliae* may originate from the disappearance of differences in trans regulators

448   between species A1 and D1 upon hybridization as in *V. longisporum* they reside in the same

449   nuclear environment (Fig 5; S2 Table) [27]. Homogenization of expression patterns of

450   homeologs has been similarly observed in the fungal allopolyploid *Epichloë* Lp1 [31].

451   Furthermore, the enrichment in genes encoding secreted proteins for *V. longisporum* genes

452   that are differently regulated upon infection indicates their importance for infection, as

453   secreted proteins play important roles in pathogen-host interactions (Gupta et al. 2015; Fig 7).

454   Upon plant infection, expression of homeologs encoding secreted proteins homogenized,

455   illustrating gene expression crosstalk between the different sub-genomes of *V. longisporum*

456   (Fig 7). Ratio changes between homeolog expressions also occurred in synthetic allopolyploid

457   *Arabidopsis* upon cold stress treatment [63]. Many of these ratio alterations were related to

458   stress responses. Thus, conceivably, alterations in homeolog expression ratios facilitate the

459   adjustment of allopolyploids to different environmental conditions.

460

461   **Conclusion**

462   Allodiploidization is an intrusive evolutionary mechanism that involves extensive alterations

463   in genome, gene and transcriptome evolution. *V. longisporum* displays signatures of rapid

464   diversifying evolution in the aftermath of hybridization, illustrated by extensive genomic

465   rearrangements and accelerated gene evolution. Furthermore, the regulatory crosstalk between

466   sub-genomes can adjust gene expression depending on the environment. Thus, in comparison

467   to non-hybrid *Verticillium* spp., *V. longisporum* has a high adaptive potential that can

468   contribute to host immunity evasion and to the further specialization towards brassicaceous

469   plant hosts.

470

471   **Material and methods**

472   **Genome analysis**

473   Genome assemblies of the two *V. longisporum* strains (VLB2 and VL20) and *V. dahliae* strain

474   JR2 were previously published [39,46]. Telomeric regions were determined based on the

475 fungal telomeric repeat pattern: TAACCC/GGGTTA (minimum three repetitions) [46].

476 Furthermore, additional repeats were identified and characterized using RepeatModeler

477 (v1.0.8). *De novo*-identified repeats were combined with the repeat library from RepBase

478 (release 20170127) [64]. The genomic coordinates of the repeats were identified with

479 RepeatMasker (v4.0.6). Homologous genes were identified by nucleotide BLAST (v2.2.31+).

480 Here, only hits with a minimal coverage of 80% with each other were selected.

481

482 **RNA sequencing, gene annotation and function determination**

483 To obtain RNA-seq data for *Verticillium* grown in culture medium, isolates JR2, VLB2 and

484 VL20 were grown for three days in potato dextrose broth (PDB) with three biological

485 replicates for every isolate. To obtain RNA-seq data from *Verticillium* grown *in planta*, two-

486 week-old plants of the susceptible oilseed rape cultivar 'Quartz' were inoculated by dipping

487 the roots for 10 minutes in $1x10^6$ conidiospores ml^{-1} spore suspension of *V. longisporum*

488 isolates VLB2 and VL20, respectively [50]. Similarly, three-week-old *A. thaliana* (Col-0)

489 plants were inoculated with *V. dahliae* isolate JR2, VLB2 and VL20, respectively. After root

490 inoculation, plants were grown in individual pots in a greenhouse under a cycle of 16 h of

491 light and 8 h of darkness, with temperatures maintained between 20 and 22°C during the day

492 and a minimum of 15°C overnight. Three pooled samples (10 plants per sample) of stem

493 fragments (3 cm) and complete flowering stems were used for total RNA extraction for

494 oilseed rape and *A. thaliana*, respectively. Total RNA was extracted based on TRIzol RNA

495 extraction (Simms et al. 1993). cDNA synthesis, library preparation (TruSeq RNA-Seq short-

496 insert library), and Illumina sequencing (single-end 50 bp) was performed at the Beijing

497 Genome Institute (BGI, Hong Kong, China). In total, ~2 Gb and ~1.5 Gb of filtered reads

498 were obtained for the *Verticillium* samples grown in culture medium and *in planta*,

499 respectively. RNAseq data were submitted to the SRA database under the accession number:

500 SRP149060.

501 Using RNA-seq from the in liquid medium grown cultures, gene annotation was

502 performed for JR2, VLB2 and VL20 with the BRAKER1 1.9 pipeline [47] using GeneMark-

503 ET [65] and AUGUSTUS [66]. Predicted genes with internal stop codons were removed from

504 the analysis. The secretome prediction was done using SingalP4 (v4.1) [67], TargetP (v1.1)

505 [68], and TMHMM (v2.0) [69] as described previously [70]. Pfam function domains were

506 predicted using InterProScan [71]. Subsequently, Pfam enrichments was determined using

507 hypergeometric tests, and significance values were corrected using the Benjamini-Hochberg

508 false discovery method [72]. Clusters of Orthologous Group (COG) categories were

509 determined for protein sequences using EggNOG (v4.5.1) [73].

510

511 **Parental origin determination**

512 Sub-genomes were divided based on the differences in sequence identities between species

513 A1 and D1 with *V. dahliae*. *V. longisporum* genomes of VLB2 and VL20 were aligned to the

514 complete genome assembly of *V. dahliae* JR2 using NUCmer, which is part of the MUMmer

515 package v3.23 [46,74]. Here, only 1-to-1 alignments longer than 10 kb and with a minimum

516 of 80% identity were retained. Subsequent alignments were concatenated if they aligned to

517 the same contig with the same orientation and order as the reference genome. The average

518 nucleotide identity was determined for every concatenated alignment and used to divide the

519 genomes into sub-genome.

520 The parental origin determination based on sequence identities of the exonic regions of

521 genes was performed by BLAST (v2.6.0+). Here, hits with a minimum subject and query

522 coverage of 80% were used. Furthermore, similar to Louis et al. (2012), differences in GC-

523 content between homologous genes present in two copies were calculated accordingly:

22

524
$$dGC_{gene} = \frac{2 * GCgene}{GCgene + GChomolog}$$

525   (1)

526   GCgene = GC% of gene

527   GChomolog = GC% of homolog

528   dGC_{gene} = GC ratio from the mean GC% value

529

530   **Gene conversion and genomic rearrangements**

531   Genes occurring in multiple copy were identified using nucleotide BLAST (v2.6.0+) and the

532   sequence identity between these genes was determined. Here, hits with a minimum subject

533   and query coverage of 80% were used.

534   The VLB2 genome assembly was aligned to VL20 to identify synteny breaks using

535   NUCmer, which is part of the MUMmer package v3.23 [74]. Subsequent alignments were

536   concatenated if they aligned to the same contig with the same orientation and order as the

537   reference genome. In order to confirm synteny breaks, filtered *V. longisporum* long

538   sequencing reads of VLB2 [39] were aligned to the *V. longisporum* VL20 genome with the

539   Burrows-Wheeler Aligner (BWA) [75] and further processed with the samtools package

540   (v1.3.1) [76]. Synteny breaks were visualized using the R package Sushi [77] and the

541   Integrative Genomics Viewer [78]. The association between breaks with repeats was tested

542   through permutation. First, the fraction of synteny breaks flanked by repeats was determined.

543   Here, synteny breaks were assigned to reside in a "repeat-rich" region if a 1 kb window

544   around the break consisted for more than 10% of repeats. The *V. longisporum* VL20 genome

545   assembly was divided into windows of 1 kb using BEDTools (v2.26.0). To estimate the

546   significance of the synteny break/repeat association [79], 10,000 permutations were executed

547   with the same amount of windows as there were synteny breaks to determine the random

548   distribution of repeat-rich regions.

23

549

Phylogenetic tree

551 Following *Verticillium* strains were used as representatives for their species: *V. albo-atrum* =

552 PD747, *V. alfalfae* = PD683, *V. dahliae* = JR2, *V. isaacii* = PD660, *V. klebahnii* = PD401, *V.*

553 *nonalfalfae* = TAB2, *V. nubilum* = PD621, *V. tricorpus* = PD593 and *V. zaregansianum* =

554 PD739. The phylogenetic tree was constructed based on nucleotide sequences of the

555 Benchmarking Universal Single-Copy Orthologs (BUSCOs) of fungi present in all

556 *Verticillium* spp. and the out-group species *Sodiomyces alkalinus* [80]. To this end, previously

557 published *Verticillium* and *S. alkalinus* assemblies were used [39,46,56,81]. In total, 277

558 orthologous groups were aligned using mafft (v7.271) (default settings) [82,83]. Aligned

559 genes were then concatenated, and the phylogenetic tree was inferred using RAxML with the

560 GTRGAMMA substitution model (v8.2.0) [84]. The robustness of the inferred phylogeny was

561 assessed by 100 rapid bootstrap approximations.

562

Gene divergence

564 Previously published annotations of the haploid *Verticillium* spp. were used to compare the

565 evolutionary speed of orthologs [46,56]. The VESPA (v1.0b) software was used to automate

566 this process [85]. The coding sequences for each *Verticillium* spp. were filtered and

567 subsequently translated using the VESPA 'clean' and 'translate' function. Homologous genes

568 were retrieved by protein BLAST (v2.2.31+) querying a database consisting of all

569 *Verticillium* protein sequences. Here, only hits with a minimum coverage of 70% were used.

570 Homologous genes were grouped with the VESPA 'best_reciprocal_group' function. Only

571 homology groups that comprised a single representative for every *Verticillium* spp. were used

572 for further analysis. Protein sequences of each homology group were aligned with muscle

573 (v3.8.31) [86]. The aligned protein sequences of the homology groups were conversed to

24

574   nucleotide sequence by the VESPA 'map_alignments' function. The alignments were used to

575   calculate *Ka/Ks* for every branch of the species phylogeny using codeml module of PAML

576   (v4.8) with the following parameters: F3X4 codon frequency model, wag.dat empirical amino

577   acid substitution model and no molecular clock [87]. To this end, the previously obtained

578   phylogenetic tree topology was used: (((((*V. klebahnii*, *V. isaacii*), *V. tricorpus*), *V.*

579   *zaregamsianum*), *V. albo-atrum*), (((*V. dahliae*, species D1), (*V. alfalfae*, *V. nonalfalfae*)),

580   species A1), *V. nubilum*). When comparing the evolutionary speed between *V. dahliae* genes

581   and its orthologs, extreme values *Ka/Ks* were discarded from further analysis, i.e. *Ka/Ks* <

582   0.0001 and *Ka/Ks* ≥ 2. Significance of positive selection was tested using a Z-test [88] and Z-

583   values >1.65 were considered significant with *P*<0.05.

584

585   **Gene expression analysis**

586   The RNA sequencing reads of the *Verticillium* strains VLB2, VL20 and JR2 were uniquely

587   mapped to their previously assembled genomes using the Rsubread package in R

588   [39,46,89,90]. To compare gene expression patterns, gene orthologs were retrieved by protein

589   BLAST (v2.2.31+). Here, only hits with a minimum sequence identity of 70% and of

590   coverage of 80% were used. Only genes in single copy in *V. dahliae* with two orthologs in *V.*

591   *longisporum* of different parental origin (one A1 copy and one D1 copy) were used for

592   comparative analysis. The comparative transcriptomic analysis was performed with the

593   package edgeR in R (v3.4.3) [90–92]. Expression patterns were corrected for putative length

594   differences between orthologous genes. Genes are considered differently expressed when *P*-

595   value < 0.05 with a log2-fold-change ≥ 1. *P*-values were corrected for multiple comparisons

596   according to Benjamini and Hochberg [72].

597

Acknowledgements

References

1.   Wolfe KH. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet. 2001;2: 333–41.

2.   Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473: 97–100.

3.   Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. Nat Rev Genet. 2017;18:411-24.

4.   Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, et al. Polyploidy and angiosperm diversification. Am J Bot. 2009;96: 336–48.

5.   Hoegg S, Brinkmann H, Taylor JS, Meyer A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol. 2004;59: 190–203.

6.   te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, et al. The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot. 2012;109: 19–45.

7.   Lohaus R, Van de Peer Y. Of dups and dinos: evolution at the K/Pg boundary. Curr Opin Plant Biol. 2016;30: 62–9.

8.   Vanneste K, Maere S, Van de Peer Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos Trans R Soc B Biol Sci. 2014;369: 20130353.

9.   Vanneste K, Baele G, Maere S, Van de Peer Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. Genome Res. 2014; 1334–47.

10.   Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. New Phytol. 2015;210: 391–8.

11.   Campbell MA, Ganley ARD, Gabaldón T, Cox MP. The case of the missing ancient fungal polyploids. Am Nat. 2016;188: 602–14.

12.   Marcet-Houben M, Gabaldón T. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. PLOS Biol. 2015;13: e1002220.

13.   Peris D, Moriarty R V, Alexander WG, Baker E, Sylvester K, Sardi M, et al. Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. Biotechnol Biofuels. 2017;10: 78.

14.   Gibson B, Liti G. *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. Yeast. 2015;32: 17–27.

15.   Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, et al. Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet. 2008;42: 443–61.

16.   Matute DR, Butler IA, Turissini DA, Coyne JA. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. Science. 2010;329: 1518–21.

17.   Mixão V, Gabaldón T. Hybridization and emergence of virulence in opportunistic human yeast pathogens. Yeast. 2017;35: 5–20.

18.   McGrath CL, Gout JF, Johri P, Doak TG, Lynch M. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. Genome Res. 2014;24: 1665–75.

19.   Louis VL, Despons L, Friedrich A, Martin T, Durrens P, Casarégola S, et al. *Pichia sorbitophila*, an interspecies yeast hybrid, reveals early steps of genome resolution after polyploidization. G3. 2012;2: 299–311.

20.   Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc Natl Acad Sci USA. 2012;109: 10954–9.

21. Ortiz-Merino RA, Kuanyshev N, Braun-Galleani S, Byrne KP, Porro D, Branduardi P, et al. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. PLOS Biol. 2017;15: e2002128.

22. Kimura M, Ohta T. On some principles governing molecular evolution. Proc Natl Acad Sci USA. 1974;71: 2848–52.

23. Hellsten U, Khokha MK, Grammer TC, Harland RM, Richardson P, Rokhsar DS. Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. BMC Biol. 2007;5: 31.

24. Sriswasdi S, Takashima M, Manabe R, Ohkuma M, Sugita T, Iwasaki W. Global deceleration of gene evolution following recent genome hybridizations in fungi. Genome Res. 2016;26: 1081–90.

25. Schranz ME, Mohammadin S, Edger PP. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr Opin Plant Biol. 2012;15: 147–53.

26. Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. 2012;196: 966–71.

27. Tirosh I, Reikhav S, Levy AA, Barkai N. A yeast hybrid provides insight into the evolution of gene expression regulation. Science. 2009;324: 659–662.

28. Adams KL, Cronn R, Percifield R, Wendel JF. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA. 2003;100: 4649–54.

29. Buggs RJA, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, et al. Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. New Phytol. 2010;186: 175–83.

30. Yoo M-J, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity. 2013;110: 171–80.

31. Cox MP, Dong T, Shen G, Dalvi Y, Scott DB, Ganley ARD. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. PLoS Genet. 2014;10: e1004180.

32. Cook DE, Mesarich CH, Thomma BPHJ. Understanding plant immunity as a surveillance system to detect invasion. Annu Rev Phytopathol. 2015;53: 541–63.

33. Jones JDG, Dangl JL. The plant immune system. Nature. 2006;444: 323–29.

34. Depotter JRL, Seidl MF, Wood TA, Thomma BPHJ. Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. Curr Opin Microbiol. 2016;32: 7–13.

35. Husson C, Aguayo J, Revellin C, Frey P, Ioos R, Marçais B. Evidence for homoploid speciation in *Phytophthora alni* supports taxonomic reclassification in this species complex. Fungal Genet Biol. 2015;77: 12–21.

36. Brasier CM, Kirk SA. Comparative aggressiveness of standard and variant hybrid alder phytophthoras, *Phytophthora cambivora* and other *Phytophthora* species on bark of *Alnus*, *Quercus* and other woody hosts. Plant Pathol. 2001;50: 218–29.

37. Inderbitzin P, Davis RM, Bostock RM, Subbarao KV. The ascomycete *Verticillium longisporum* is a hybrid and a plant pathogen with an expanded host range. PLoS One. 2011;6: e18260.

38. Zeise K, von Tiedemann A. Host specialization among vegetative compatibility groups of Verticillium dahliae in relation to *Verticillium longisporum*. J Phytopathol. 2002;150: 112–119.

39. Depotter JRL, Seidl MF, van den Berg GCM, Thomma BPHJ, Wood TA. A distinct and genetically diverse lineage of the hybrid fungal pathogen *Verticillium longisporum*

population causes stem striping in British oilseed rape. Environ Microbiol. 2017;19: 3997–4009.

40. Eynck C, Koopmann B, Grunewaldt-Stoecker G, Karlovsky P, von Tiedemann A. Differential interactions of *Verticillium longisporum* and *V. dahliae* with *Brassica napus* detected with molecular and histological techniques. Eur J Plant Pathol. 2007;118: 259–74.

41. Ellendorff U, Fradin EF, de Jonge R, Thomma BPHJ. RNA silencing is required for *Arabidopsis* defence against Verticillium wilt disease. J Exp Bot. 2009;60: 591–602.

42. Fradin EF, Abd-El-Haliem A, Masini L, van den Berg GCM, Joosten MHAJ, Thomma BPHJ. Interfamily transfer of tomato *Ve1* mediates *Verticillium* resistance in *Arabidopsis*. Plant Physiol. 2011;156: 2255–65.

43. Short DPG, Gurung S, Hu X, Inderbitzin P, Subbarao KV. Maintenance of sex-related genes and the co-occurrence of both mating types in *Verticillium dahliae*. PLoS One. 2014;9: e112145.

44. Novakazi F, Inderbitzin P, Sandoya G, Hayes RJ, von Tiedemann A, Subbarao KV. The three lineages of the diploid hybrid *Verticillium longisporum* differ in virulence and pathogenicity. Phytopathology. 2015;105: 662–73.

45. Depotter JRL, Deketelaere S, Inderbitzin P, von Tiedemann A, Höfte M, Subbarao KV, et al. *Verticillium longisporum* , the invisible threat to oilseed rape and other brassicaceous plant hosts. Mol Plant Pathol. 2016;17: 1004–16.

46. Faino L, Seidl M, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, et al. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. MBio. 2015;6: e00936-15.

47. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS.

Bioinformatics. 2016;32: 767–9.

48.   Seidl MF, Thomma BPHJ. Sex or no sex: evolutionary adaptation occurs regardless. BioEssays. 2014;36: 335–45.

49.   Gupta R, Lee SE, Agrawal GK, Rakwal R, Park S, Wang Y, et al. Understanding the plant-pathogen interactions in the context of proteomics-generated apoplastic proteins inventory. Front Plant Sci. 2015;6: 352.

50.   Depotter JRL, Rodriguez-Moreno L, Thomma BPHJ, Wood TA. The emerging British *Verticillium longisporum* population consists of aggressive *Brassica* pathogens. Phytopathology. 2017;107: 1399–1405.

51.   Stergiopoulos I, Kourmpetis YAI, Slot JC, Bakker FT, De Wit PJGM, Rokas A. In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. Mol Biol Evol. 2012;29: 3371–84.

52.   Marín-Rodríguez MC, Orchard J, Seymour GB. Pectate lyases, cell wall degradation and fruit softening. J Exp Bot. 2002;53: 2115–9.

53.   Stukenbrock EH. The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology. 2016;106: 104–12.

54.   de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 2013;23: 1271–82.

55.   Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 2016;26: 1091–100.

56.   Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF. Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangements and gene losses. Environ Microbiol. 2018; 20:1362-73.

57.   Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2001;290: 1151–5.

58.   McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature. 1991;351: 652–4.

59.   Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, et al. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA. 2005;102: 5454–9.

60.   Dutheil JY, Mannhaupt G, Schweizer G, Sieber CMK, Münsterkötter M, Güldener U, et al. A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. Genome Biol Evol. 2016;8: 681–704.

61.   Fouche S, Plissonneau C, Croll D. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr Opin Microbiol. 2018;46: 34–42.

62.   Shi X, Ng DWK, Zhang C, Comai L, Ye W, Chen ZJ. Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nat Commun. 2012;3: 950.

63.   Akama S, Shimizu-Inatsugi R, Shimizu KK, Sese J. Genome-wide quantification of homeolog expression ratio revealed nonstachastic gene regulation in synthetic allopolyploid *Arabidopsis*. Nucleic Acids Res. 2014;42: e46.

64.   Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6: 11.

65.   Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 2014;42: e119.

66.   Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24: 637–44.

67. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8: 785–6.

68. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 2000;300: 1005–16.

69. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol. 2001;305: 567–80.

70. Seidl MF, Faino L, Shi-Kunne X, van den Berg GCM, Bolton MD, Thomma BPHJ. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. Mol Plant-Microbe Interact. 2015;28: 362–73.

71. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30: 1236–40.

72. Benjamini Y, Hochberg Y. Controlling the false discovery rate : a practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57: 289–300.

73. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44: D286–D293.

74. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5: R12.

75. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. Preprint. Available from: arXiv: 1303.3997. Cited 28 May 2018.

76. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence

Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–9.

77.   Phanstiel DH, Boyle AP, Araya CL, Snyder MP. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics. 2014;30: 2808–10.

78.   Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. Nat Biotechnol. 2011;29: 24–6.

79.   Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–2.

80.   Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV., Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31: 3210–2.

81.   Grum-Grzhimaylo AA, Debets AJM, van Diepeningen AD, Georgieva ML, Bilanenko EN. *Sodiomyces alkalinus*, a new holomorphic alkaliphilic ascomycete within the Plectosphaerellaceae. Persoonia. 2013;31: 147–58.

82.   Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30: 772–80.

83.   Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30: 3059–66.

84.   Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–13.

85.   Webb AE, Walsh TA, O'Connell MJ. VESPA: Very large-scale Evolutionary and Selective Pressure Analyses. PeerJ Comput Sci. 2017;3: e118.

86.   Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32: 1792–7.

87.   Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24: 1586–91.

88.   Stukenbrock EH, Dutheil JY. Comparing fungal genomes: insight into functional and evolutionary processes. Methods Mol Biol. 2012;835: 531–48.

89.   Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41: e108.

90.   R Core Team. [cited 28 May 2018] R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: https://www.r-project.org/

91.   Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–40.

92.   McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40: 4288–97.

93.   Mccarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics. 2009;25: 765–71.

Supporting informations

S1 Table. Comparison *V. longisporum* and *V. dahliae* genomes.

S2 Table. Expression pattern correlation of genes between *V. dahliae* and *V. longisporum* sub-genomes. VD = *V. dahliae*, A1 = *V. longisporum* A1 sub-genome and D1 = *V. longisporum* D1 sub-genome. Correlations are calculated with the Spearman's rank correlation coefficient based on the count per million (CPM) values.

S3 Table. Pfam domain enrichment for differentially expressed *V. longisporum* (expression A1 and D1 homeologs combined) and *V. dahliae* orthologs in culture medium

S1 Fig. Lines of evidence for the parental origin of *V. longisporum* genomic regions. (A) Distribution of sequence identity of *V. longisporum* alignments to *V. dahliae*. (B) The distribution of the sequence identity between *V. longisporum* exonic regions of genes and their *V. dahliae* orthologs. (C) Distribution of sequence identity between exonic regions of *V. longisporum* homologs that are present in two copies. Strains VLB2 and JR2 were used for *V. longisporum* and *V. dahliae*, respectively.

S2 Fig. Gene copy number distribution within *Verticillium* (sub-)genomes. "(A1)" and "(D1)" represent species A1 and D1 sub-genomes, respectively, of the *V. longisporum* strains VLB2 and VL20. For *V. dahliae*, the strain JR2 was used.

37

Fig S3: Synteny break confirmation by mapping *V. longisporum* VLB2 reads to the VL20 genome. Red and blue bars represent forward and reverse aligned VLB2 reads to the VL20 genome. The dashed lines show the suggested position in synteny break through genome alignment. (A) In particular cases, read alignments did not confirm the break in synteny. (B) In other cases, breaks in synteny were confirmed as reads abruptly stopped and started on these genome positions. (C) Breaks were also considered truthful if regions showed overlap in repeat-rich regions where read overlap between adjacent genome regions is lacking.

Fig S4: The association of synteny breaks with repetitive elements. The black curve represents the fraction of 60 randomly chosen 1 kb windows in the *V. longisporum* VL20 that are repeat-rich, which has been permutated 10,000 times. The red line indicates the fraction of true breaks that lay in a 1 kb window enriched for repeats (57%).

Fig S5: Gene evolutionary speed comparison between *Verticillium* species. *Ka/Ks* ratios (ω) between *V. dahliae* and other *Verticillium* spp. were compared. (A) Blue branches in the phylogenetic tree represent *Verticillium* species with a haploid evolutionary history, whereas red branches represent species with an allodiploid phase in their evolutionary history. The median difference in evolutionary speed between *V. dahliae* and orthologs was calculated in addition to the standard deviation (σ). (B) The distribution of these differences in evolutionary speed is depicted for species of the *Verticillium* clade Flavnonexudans. For the *Verticillium* species A1 and D1, *V. longisporum* strain VL20 was used.

Fig S6: Genes under positive selection in *Verticillium* species. Genes with one ortholog in every *Verticillium* spp. were tested for positive selection using a *Z*-test ($P<0.05$). # = number of genes under positive selection SP = number of genes encoding secreted proteins under positive selection.

Fig S7: Relative abundance of Clusters of Orthologous Group (COG) categories for genes under positive selection. Only genes were included were the COG category/categories could be determined. Genes under positive selection had a *Z*-value > 1.65 ($P<0.05$). For the *Verticillium* species A1 and D1, *V. longisporum* strain VLB2 was used.

Fig S8: Differential expression in *V. longisporum* VL20 in relation to genes encoding secreted proteins. Differential expression is calculated for species A1, species D1 and *V. longisporum* (expression A1 and D1 homeologs combined) with *V. dahliae* orthologs. Pie charts show fraction of genes encoding secreted proteins of not-differentially and differentially expressed genes, respectively. Significance of the different distributions was calculated the Fisher's exact test (ns = not significant, *** = $P<0.001$).

Fig S9: Expression comparison of genes with a pectate lyase domain (PF03211). Gene expressions are depicted for species A1 and species D1 homeologs of *V. longisporum* cultured in liquid medium and upon oilseed rape colonization, respectively. Bars represent the mean gene expression and error bars represent the standard deviation. The significance of difference in gene expression was calculated using *t*-tests relative to a threshold (TREAT) [93]. Level of significance: ns = not significant and *** = $P<0.001$.

FIGURE CAPTIONS

Fig 1. Determination of parental origin of *Verticillium longisporum* genome sections. The ten largest contigs of the genome assemblies of *V. longisporum* strains VLB2 and VL20 are depicted. Lane 1: regions with a sequence identity of at least 96% were assigned to Species D1 (orange), whereas ones with lower sequence identity to Species A1 (purple). Lane 2: sequence identity of *V. longisporum* alignments to *V. dahliae*. Lane 3: Sequence identity between exonic regions of *V. longisporum* and *V. dahliae* orthologs. Lane 4: Difference in sequence identity in percent point (pp) between exonic regions of *V. longisporum* double-copy genes. Only gene pairs with an ortholog in *V. dahliae* are depicted. Alleles with a higher identity to *V. dahliae* are depicted as a positive pp difference, whereas the corresponding homolog as a negative pp difference. Lane 5: the relative difference in GC content (dGC) between genes in double copy. Lane 6: Read depth with non-overlapping windows of 10 kb. Data points of lanes 3-5 represent the average value of a window of eleven genes, which proceed with a step of one gene.

Fig 2. The origin of the mosaic genome structure of *V. longisporum*. (A) The contribution of gene conversion to *V. longisporum* genome evolution. Sequence identities between genes in copy, present in *V. longisporum* VLB2 and VL20, are depicted. Gene pairs that encountered gene conversion (purple dots in the red zones) have sequence divergence of more than one percent in one *V. longisporum* strain and less than one percent in the other strain. In other cases, pairs that differ less than one percent are depicted as a black dot, whereas a difference greater than one percent is depicted as a blue dot. (B) The contribution of genomic rearrangements to *V. longisporum* genome evolution. The ten largest contigs of the *V. longisporum* strains VLB2 (displayed on the right) and VL20 (displayed on the left) are depicted with complete chromosomes indicated by asterisks. Ribbons indicate syntenic genome regions between the two strains and contig colors indicate the parental origin similar to Fig 1 (purple = A1 and orange = D1). Regions without ribbons do not have homology to regions in the 10 largest contigs of the other isolate. Red bars on the contigs indicate synteny breaks that are confirmed by discontinuity in read alignment of VLB2 to the VL20 genome assembly (S3 Fig).

Fig 3. Comparison of gene divergence between *Verticillium* species. *Ka/Ks* ratios (ω) between *V. dahliae* and other *Verticillium* spp. were compared. (A) Blue branches in the phylogenetic tree represent *Verticillium* species with a haploid evolutionary history, whereas red branches represent species with an allodiploid phase in their evolutionary history. The median difference in evolutionary speed between *V. dahliae* and orthologs was calculated in addition to the standard deviation (σ). (B) The distribution of these differences is depicted for species of the *Verticillium* clade Flavnonexudans. For the *Verticillium* species A1 and D1, *V. longisporum* strain VLB2 was used.

Fig 4. Global comparison of expression between *V. longisporum* and *V. dahliae* orthologs. Only genes with one copy in *V. dahliae* and two orthologs in the *V. longisporum* strains VLB2 and VL20 (one from sub-genome A1 and one from sub-genome D1) were considered.

Fig 5. Expression pattern comparison between *V. longisporum* sub-genomes and *V. dahliae*. Only genes with one copy in *V. dahliae* and two orthologs in the *V. longisporum* strains VLB2 and VL20 (one from sub-genome A1 and one from sub-genome D1) were considered. (A) Pairwise correlations between *Verticillium* expression patterns. Dot size and colour represent the Spearman's correlation coefficient (ρ). (B) Relative expression (Z-scores) between *Verticillium* spp. VD = *V. dahliae*, A1 = *V. longisporum* A1 sub-genome and D1 = *V. longisporum* D1 sub-genome.

Fig 6. Differential expression between *V. longisporum* and *V. dahliae* in relation to genes encoding secreted proteins. Differential expression is calculated for species A1, species D1 and *V. longisporum* (cumulative expression A1 and D1 homeologs) with *V. dahliae* orthologs for isolates grown in (A) culture medium and (B) *in planta*. Pie charts show fraction of genes encoding secreted proteins of not-differentially and differentially expressed genes, respectively. Significance of the different distributions was calculated the Fisher's exact test (*** = $P<0.001$).

Fig 7: Differential expression between *V. longisporum* sub-genomes in relation to genes encoding secreted proteins. Gene expression of A1 and D1 homeologs were compared for *V. longisporum* grown in culture medium and *in planta*. Only genes were considered with one copy in *V. dahliae* and two corresponding copies in the *V. longisporum* strains VLB2 (one from sub-genome A1 and one from sub-genome D1). The number below the bar indicates the amount of differently and not differently expressed genes, respectively. Significance of the different distributions was calculated the Fisher's exact test (ns = not significant, * = *P<0.05*, *** = *P<0.001*).
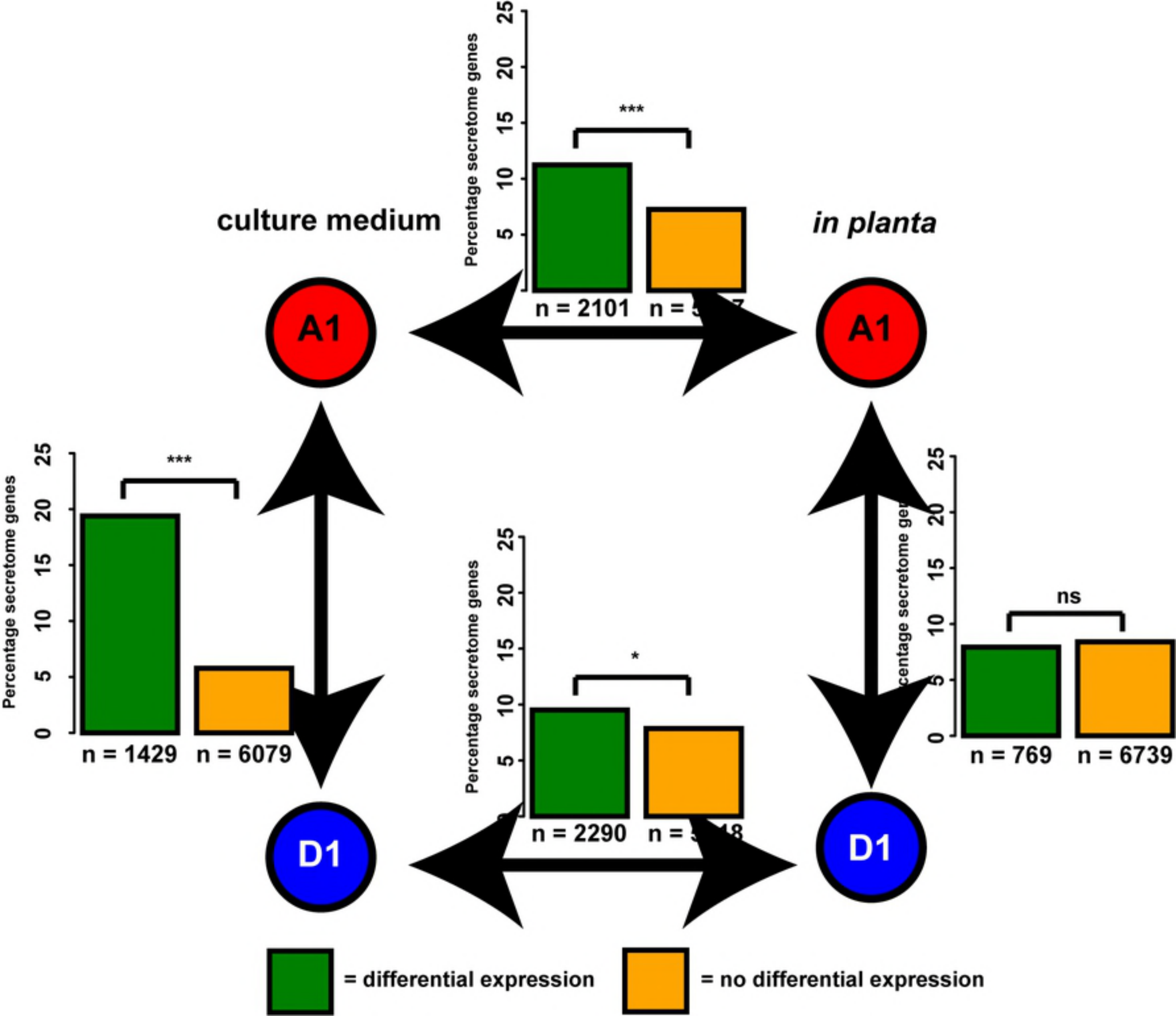
Lanes

1) Parental origin genome regions
2) Alignment sequence identity to *V. dahliae* (%)
3) Sequence identity to *V. dahliae* orthologs (%)
4) Sequence divergence between double-copy genes (pp)
5) Differential GC content between double-copy genes (dGC)
6) Read depth
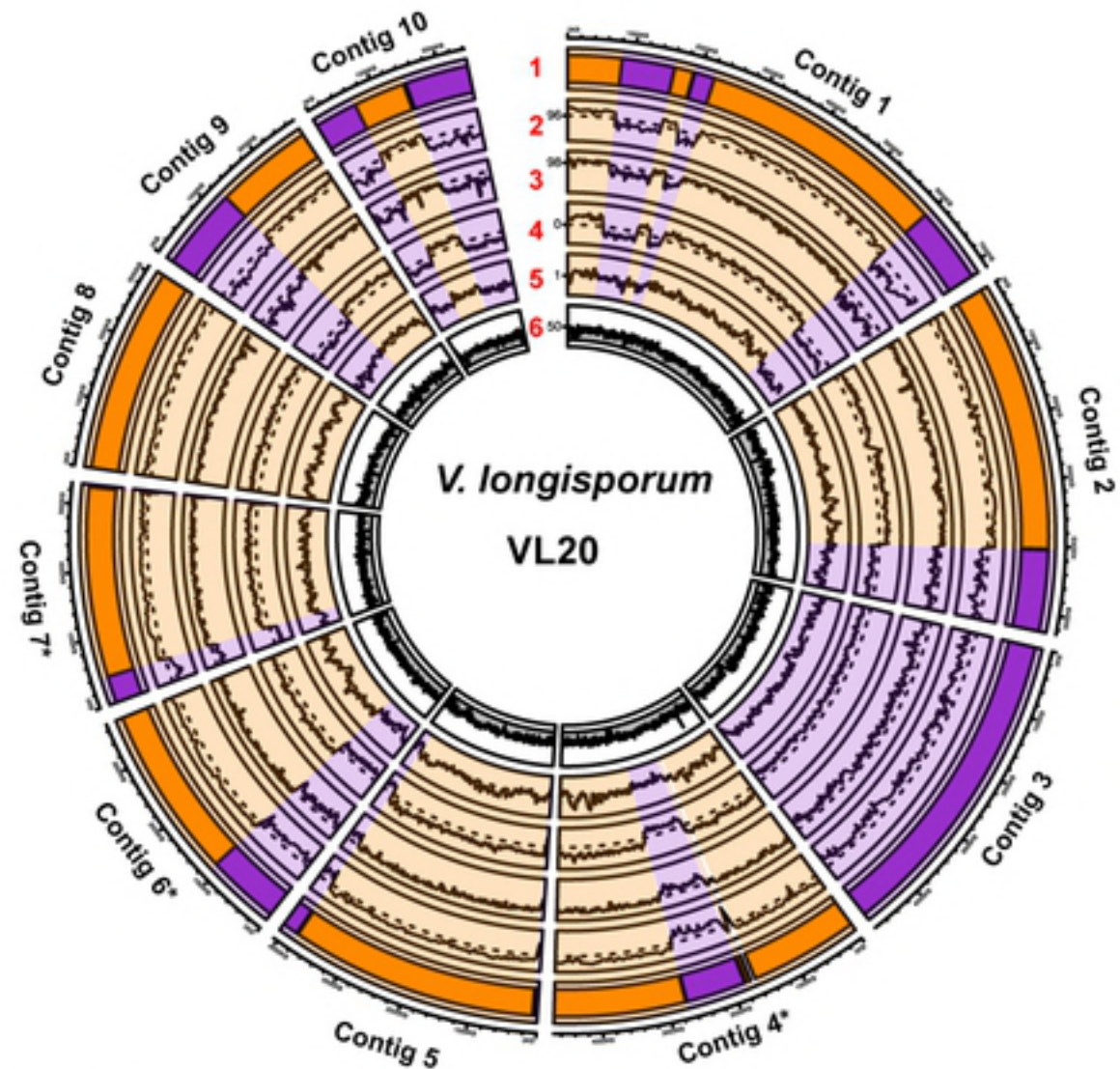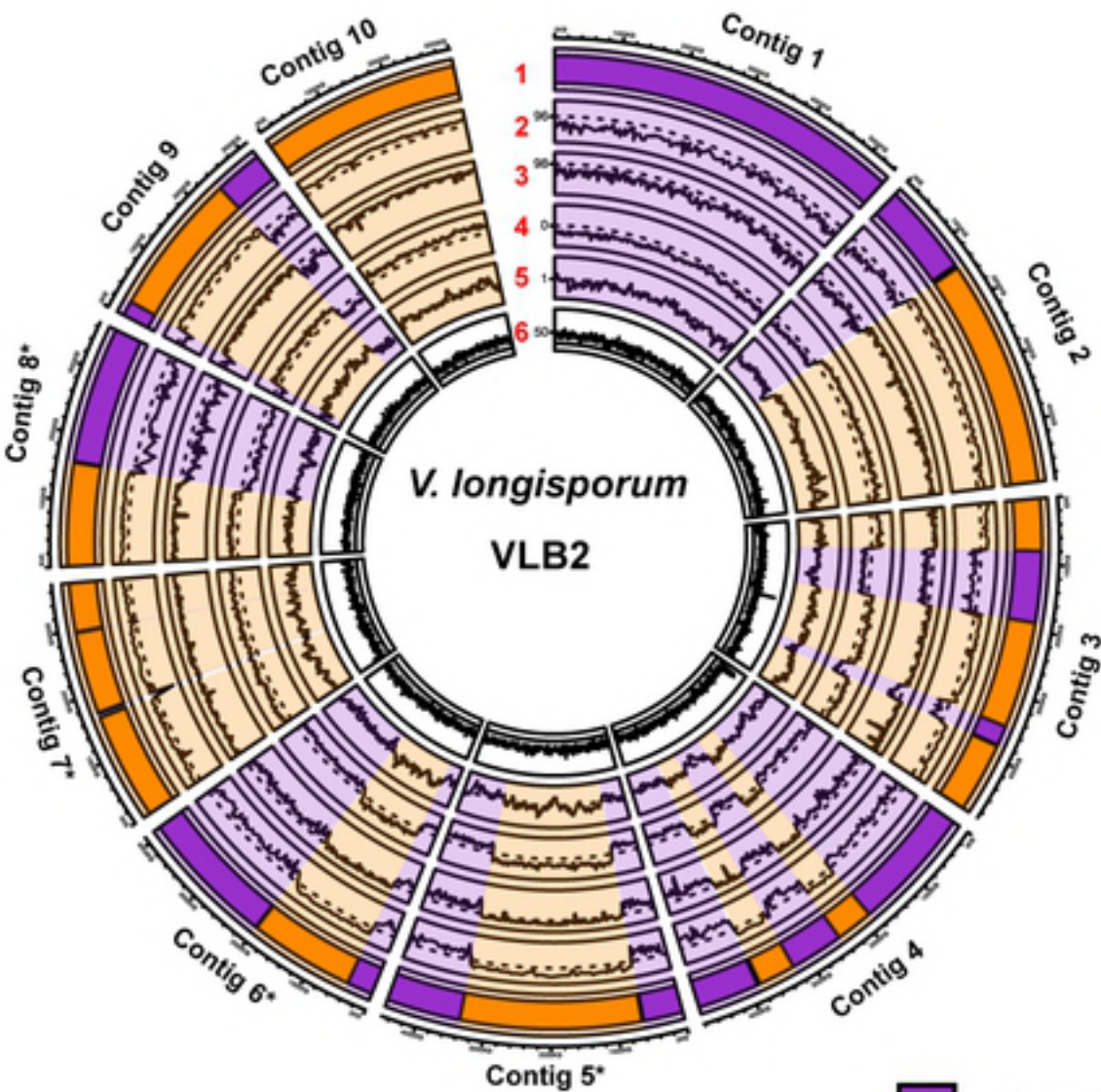
= Species A1 sub-genome

= Species D1 sub-genome

A
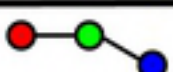
density

Sequence identity homologs VL20 (%)

Sequence identity homologs VLB2 (%)

density

B

VL20   Con 7* Con 8 Con 9 Con 10 Con 1   VLB2

Con 6*

Con 5

Con 4*

Con 3

Con 2

Con 1

Con 10 Con 9 Con 8* Con 7* Con 6* Con 5* Con 4 Con 3 Con 2

A

	$\log_2(\omega_{species}/\omega_{V.\,dahliae})$	
	median	σ
V. nubilum	-1.80	1.52
A1	-0.52	1.66
V. alfalfae	-0.88	1.61
V. nonalfalfae	-0.65	1.63
V. dahliae	0.00	0.00
D1	0.27	1.63
V. albo-atrum	-1.84	1.54
V. zaregansianum	-1.24	1.69
V. tricorpus	-1.57	1.60
V. isaacii	-1.40	1.68
V. klebahnii	-1.22	1.74

Clade Flavnonexudans

Clade Flavexudans

0.07

B

Slower than *V. dahliae* ortholog

Faster than *V. dahliae* ortholog

- species A1
- species D1
- *V. alfalfae*
- *V. nonalfalfae*
- *V.nubilum*

Fraction

$\log_2(\omega_{species}/\omega_{V.\,dahliae})$

A

B

Relative expression

genes

correlation coefficient

Row Z-Score

Frequency

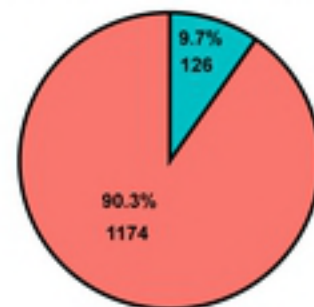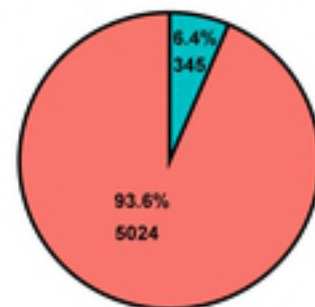**A** **Culture medium**

B *In planta*

***

No differential expression | Differential expression

No differential expression | Differential expression

5.3% 294 / 94.7% 5222

16.5% 323 / 83.5% 1630

6.4% 345 / 93.6% 5024

9.7% 126 / 90.3% 1174

= no secreted protein
= secreted protein

***

species A1

species D1

species A1

species D1

***

***

***

***

No differential expression | Differential expression
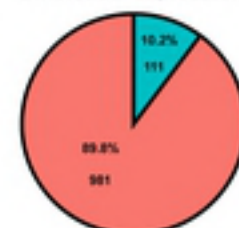
No differential expression | Differential expression

No differential expression | Differential expression

No differential expression | Differential expression

5.2% 272 / 94.8% 4924

15.2% 345 / 84.8% 1928

6.2% 350 / 93.8% 5270

14.4% 267 / 85.6% 1582

6.5% 360 / 93.5% 5217

10.2% 111 / 89.8% 981

6.6% 371 / 93.4% 5293

10.0% 100 / 90.0% 906