A distinct and genetically diverse lineage of the hybrid fungal pathogen *Verticillium longisporum* population causes stem striping in British oilseed rape

Jasper R. L. Depottet,1,2† Michael F. Seidl,1† Grardy C. M. van den Berg,1 Bart P. H. J. Thomma1,4*, and Thomas A. Wood2*,
1Laboratory of Phytopathology, Wageningen University and Research, Droevendaalsesteeg 1, Wageningen 6708 PB, The Netherlands.
2Department of Crops and Agronomy, National Institute of Agricultural Botany, Huntingdon Road, Cambridge CB3 0LE, UK.

**Summary**

Population genetic structures illustrate evolutionary trajectories of organisms adapting to differential environmental conditions. *Verticillium* stem striping disease on oilseed rape was mainly observed in continental Europe, but has recently emerged in the United Kingdom. The disease is caused by the hybrid fungal species *Verticillium longisporum* that originates from at least three separate hybridization events, yet hybrids between *Verticillium* progenitor species A1 and D1 are mainly responsible for *Verticillium* stem striping. We reveal a hitherto un-described dichotomy within *V. longisporum* lineage A1/D1 that correlates with the geographic distribution of the isolates with an ‘A1/D1 West’ and an ‘A1/D1 East’ cluster. Genome comparison between representatives of the A1/D1 West and East clusters excluded population distinctiveness through separate hybridization events. Remarkably, the A1/D1 West population that is genetically more diverse than the entire A1/D1 East cluster caused the sudden emergence of *Verticillium* stem striping in the UK, whereas in continental Europe *Verticillium* stem striping is predominantly caused by the more genetically uniform A1/D1 East population. The observed genetic diversity of the A1/D1 West population argues against a recent introduction of the pathogen into the UK, but rather suggests that the pathogen previously established in the UK and remained latent or unnoticed as oilseed rape pathogen until recently.

**Introduction**

*Verticillium* species are causal agents of wilt diseases on many economically important crops, with a total estimated annual loss of €3 billion worldwide in the 20 most affected hosts, including cotton and olive (Depottet *et al.*, 2016a). *Verticillium dahliae* is the most notorious wilt agent of this genus and is characterized by its extremely broad host range that encompasses hundreds of hosts (Fradin and Thomma, 2006; Inderbitzin and Subbarao, 2014). *V. dahliae* propagates mainly asexually and genomic variation is mostly established by mechanisms different from meiotic recombination, such as large-scale genomic rearrangements, horizontal gene transfer and transposon activity (de Jonge *et al.*, 2012; 2013; Seidl and Thomma, 2014; Faino *et al.*, 2016). Moreover, particular *Verticillium* spp. experienced more intrusive genomic evolution by inter-specific crosses within the genus leading to an approximate doubling of the genome size. Interspecific *Verticillium* hybrids gave rise to new diseases such as *Verticillium* stem striping on oilseed rape (Inderbitzin *et al.*, 2011b; Depottet *et al.*, 2016a). At least three hybridization events between two separate *Verticillium* spp. have occurred that have been classified under the same species name, *V. longisporum* (Karapapa *et al.*, 1997; Inderbitzin *et al.*, 2011b). The three hybrid lineages have been named after their respective hybridization parents: A1/D1, A1/D2 and A1/D3 (Inderbitzin *et al.*, 2011b). A1 and D1 are hitherto undescribed *Verticillium* species, whereas D2 and D3 are presumed *V. dahliae* isolates. Similar to other hybrid pathogens, hybridization appears to have altered the host range of *Verticillium* (Zeise and Tiedemann, 2002; Depottet *et al.*, 2016b). *V. longisporum* is highly adapted to brassicaceous hosts, such as oilseed rape and cauliflower, whereas *V. dahliae* generally does not colonize these plants (Inderbitzin and Subbarao, 2014; Depottet *et al.*, 2016a). Moreover, differences in pathogenicity are also observed between hybrid lineages, as *V. longisporum* A1/
D1 and A1/D3 are often found on multiple brassicaceous hosts, whereas lineage A1/D2 has hitherto only been found on horseradish (Inderbitzin et al., 2011b; Yu et al., 2016). Furthermore, lineage A1/D1 is predominantly found on oilseed rape and responsible for the Verticillium stem striping disease as this lineage is the most virulent on this crop (Novakazi et al., 2015).

In the past, *V. longisporum* and *V. dahliae* were considered a single species, and only at the end of the 1990s *V. longisporum* was proposed as a separate species from *V. dahliae* (Karapapa et al., 1997). While adequate characterization of *V. longisporum* strains in predating publications is obscured by this name change and taxonomic disagreements, Verticillium stem striping is exclusively caused by *V. longisporum* (Eynck et al., 2007). This oilseed rape disease was first reported in the west and south of Scania, southern Sweden, in 1969 (Kroeker, 1970). Until recently, Verticillium stem striping was only present in north-central Europe, but over the last decade other important oilseed rape production regions have also been affected (Gladders et al., 2011; CFIA, 2015). Verticillium stem striping was noticed in UK oilseed rape production for the first time in 2007, in the counties Kent and Herefordshire (Gladders et al., 2011). Intriguingly, the most likely first description of *V. longisporum* infection was in the UK during the 1950s, yet on a different host: Brussels sprout (Isaac, 1957). Currently, *V. longisporum* is present throughout England, yet the disease is most prevalent in the east (Gladders et al., 2013). The main causal agent of Verticillium stem striping, lineage A1/D1, has also been found outside Europe in Japan and the USA, albeit on different crops than oilseed rape (Carder and Barbara, 1994; Subbarao et al., 1995; Heale and Karapapa, 1999). The wide-spread occurrence of *V. longisporum* A1/D1 suggests the importance of human activity in the spread of this disease as, similar to other *Verticillium* spp. (Atallah et al., 2012), *V. longisporum* is considered soil-borne without the long-distance dispersal of air-borne spores (Depotter et al., 2016a). Dispersal of *Verticillium* by the trade of plant commodities has been observed for *V. dahliae*, which has facilitated the intercontinental spread of the pathogen (Atallah et al., 2012).

The aim of this study is to reveal population structures within Verticillium stem striping lineage A1/D1 by screening of rapidly evolving DNA regions. Microsatellites or simple sequence repeat (SSR) loci are hyper-variable genome regions of simple DNA motifs repeated in tandem. The repetitive character of these regions makes them more prone to mutation than non-repetitive sequences due to unequal crossing-over and replication slippage, generally revealing high degrees of polymorphism (Levinson and Gutman, 1987). Several population studies previously used SSR loci to describe diversity within *Verticillium* populations (Atallah et al., 2010; 2012; Short et al., 2014). However, hitherto, no population studies have been performed on *V. longisporum*. Here, we assessed the genetic diversity within a broad geographic range of the *V. longisporum* lineage A1/D1 isolates and found clear population structuring. The origin of these distinct populations was further elucidated by genealogical analysis and genome comparison. Here, established Verticillium stem striping populations (e.g. from Germany and Sweden) were compared with populations from a recent disease outbreak in the UK, in order to link the population dynamics with the expansion of Verticillium stem striping.

**Results**

**Population structure analysis**

A geographically diverse collection of *V. longisporum* isolates was obtained in order to assess the population diversity and to genotypically link isolates from different origins. In total, 88 *V. longisporum* isolates from 9 different countries were screened for SSR polymorphisms in order to determine the population structure (Supporting Information Table S1). Sixty one putative polymorphic SSR markers were developed based on unpublished draft *V. longisporum* genome sequence data and tested on the *V. longisporum* collection, of which 9 displayed polymorphisms (Table 1). Our analysis was performed as for a haploid organism as all markers only gave a single polymorphic signal.

The population structure was assessed based on the polymorphisms of these nine SSR loci that are dispersed over the genome (Table 1). The acquired multi-locus genotype (MLG) data were used to determine the most likely amount of genetic clusters in the *V. longisporum* population, allowing individual isolates to admix between different genetic clusters. The ad-hoc statistic $\Delta K$ was maximized for 3 genetic clusters in the population ($K = 3$), indicating that 3 is the most likely number of genetic clusters for the complete data set (Fig. 1; Supporting Information Fig. S1). Four isolates that were previously determined to belong to lineage A1/D3 (Inderbitzin et al., 2011b) formed one genetic cluster, whereas isolates known to belong to lineage A1/D1 showed two distinct clusters: one that includes samples from the USA and Japan, and another with samples from Germany and Sweden (Fig. 1 and Supporting Information Fig. S1). The hitherto uncharacterized isolates all grouped together with one of these two A1/D1 clusters, indicating that they belong to the A1/D1 lineage (Fig. 1 and Supporting Information Fig. S1). Indeed, successful amplification of the lineage A1/D1 specific primers D1f/AH1D1r (Inderbitzin et al., 2013) confirmed that these isolates belong to the A1/D1 lineage (Supporting Information Fig. S2). In addition, similar to all tested *V. longisporum* isolates, all uncharacterized isolates contained *MAT1–1* idiomorphs and failed to display *MAT1–2* in a mating type.
idiomorph PCR screen (Inderbitzin et al., 2011b) (Supporting Information Fig. S3). Interestingly, the dichotomy within the A1/D1 lineage correlated with the country of origin of the isolates ($p = -0.51$, $p = 0.00$), as the Belgian, Dutch, UK and USA isolates formed one of the A1/D1 clusters (A1/D1 West; $n = 44$), whereas the German, Latvian and Table 1. Polymorphic simple sequence repeat markers for Verticillium longisporum lineages A1/D1 and A1/D3.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Size (bp)</th>
<th>GenBank accession#</th>
<th>Marker location</th>
<th>Repeat motif</th>
<th>Allele#</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSR25</td>
<td>283</td>
<td>KY828946</td>
<td>6</td>
<td>(TGCC)$_{13}$</td>
<td>6</td>
<td>F: AAGGAAACCCCAATGCAACCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: GGCACTTGCCGTAGGTTAGGTA</td>
</tr>
<tr>
<td>SSR15</td>
<td>318</td>
<td>KY828947</td>
<td>3</td>
<td>(CGT)$_{7}$</td>
<td>2</td>
<td>F: GTCGATTTGCGTTTGGGAAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TACACAGTCAACGACGAGG</td>
</tr>
<tr>
<td>SSR159f</td>
<td>212</td>
<td>KY828948</td>
<td>11</td>
<td>(TCT)$_{12}$</td>
<td>10</td>
<td>F: ATCCACCATGCTAAACCGTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: ACGAATGGAAGGAAACAC</td>
</tr>
<tr>
<td>SSR27</td>
<td>294</td>
<td>KY828949</td>
<td>13</td>
<td>(GTCA)$_{17}$</td>
<td>3</td>
<td>F: CTTCTGTGTTTTGCGGAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: CACTGTTCACCGACACCC</td>
</tr>
<tr>
<td>SSR135</td>
<td>226</td>
<td>KY828950</td>
<td>30</td>
<td>(AGC)$_{11}$</td>
<td>2</td>
<td>F: ACCATGCTCTCTCGGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: AGGTCCCTGAGAAAGCCACT</td>
</tr>
<tr>
<td>SSR70</td>
<td>252</td>
<td>KY828951</td>
<td>14</td>
<td>(CA)$_{15}$</td>
<td>6</td>
<td>F: GCCAGTGCCTCTTCCTCTAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: GACCGAACACGGAGATGAECT</td>
</tr>
<tr>
<td>SSR84</td>
<td>272</td>
<td>KY828952</td>
<td>3</td>
<td>(CAG)$_{9}$</td>
<td>4</td>
<td>F: TGATTAACCTGGAAGACGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: GCCAGAAGAAACGGAGTACCT</td>
</tr>
<tr>
<td>SSR101</td>
<td>363</td>
<td>KY828953</td>
<td>12</td>
<td>TGGTCATGTCCT</td>
<td>2</td>
<td>F: TGGATCTAGTATGGAGGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TGGCTGACGATTCACCT</td>
</tr>
<tr>
<td>SSR219</td>
<td>477</td>
<td>KY828954</td>
<td>13</td>
<td>(CA)$<em>{13}$(A)$</em>{8}$</td>
<td>3</td>
<td>F: ACAACACCTGCTAATGCCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: GCAAAACATTGTTAACGCGCT</td>
</tr>
</tbody>
</table>

a. Amplicon length for isolate VLB2.
b. Amplified sequence for isolate VLB2, acquired from VLB2 genome sequence.
c. Contig number of the VLB2 genome sequence (Fig. 3).
d. Number of repeats identified in genome sequence VLB2.
e. Numbers of alleles observed in tested population.

Fig. 1. Clustering of individual Verticillium longisporum multi-locus genotypes (MLGs) using 15 polymorphic simple sequence repeat markers. Population clustering was executed on the whole data set for (A) three genetic clusters ($K = 3$) and (B) four genetic clusters ($K = 4$) with by the software Structure version 2.3 (Pritchard et al., 2000). The thick vertical bars separate the MLGs by country of origin. The bar width of every country is relative to the amount of samples: Belgium ($n = 5$), France ($n = 6$), Germany ($n = 19$), Japan ($n = 5$), Latvia ($n = 1$), the Netherlands ($n = 2$), Sweden ($n = 15$), UK ($n = 25$), USA ($n = 6$) and the cluster with isolates from the A1/D3 lineage ($n = 4$). The different colours represent separate genetic clusters. In A: red = lineage A1/D1 West, green = lineage A1/D1 East, and blue = lineage A1/D3. Lineage A1/D1 West is subdivided in section B, where yellow = A1/D1 West cluster 1, and orange = A1/D1 West cluster 2.

© 2017 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*, 00, 00–00
Swedish formed the other cluster (A1/D1 East; n = 40). These two clusters within lineage A1/D1 are further referred to as 'lineage A1/D1 West' and 'lineage A1/D1 East' according to their relative geographic location in Europe. Furthermore, isolates from both genetic clusters are found in France and in Japan.

In order to confirm the dichotomy within lineage A1/D1 and to reveal more potential sub-structuring in the population, six additional published polymorphic SSR loci were used in the analysis (Table 2) (Barbara et al., 2005; Atallah et al., 2010). The SSR marker VDA817 from Barbara et al. (2005) also revealed polymorphisms, but was targeting the same SSR locus as VDA873. Thus, VDA817 was excluded from the population analysis. In total, 13 SSR loci were polymorphic for lineage A1/D1 as two of the nine previously mentioned SSR loci (SSR101 and SSR135) only differentiated between the A1/D1 and A1/D3 lineages (Table 1). The dichotomy within lineage A1/D1 was confirmed with these additional SSR loci, as the most likely number of genetic clusters was two for the A1/D1 population (Fig. 1A; Supporting Information Fig. S3). Here, nine of the 13 polymorphic SSR loci for lineage A1/D1 subdivided the population significantly into the A1/D1 West and East dichotomy based on the analysis of molecular variance (AMOVA) ($\Phi_{ST}$ statistic, Table 2). Moreover, isolates of A1/D1 West and A1/D1 East lineages have no alleles in common for the loci SSR25, SSR70, SSR219, VD8 and VDA823. All SSR loci combined, over 59% of the total genotypic variability within the A1/D1 population can be explained by the A1/D1 West and East dichotomy. Based on this, the dichotomy within A1/D1 is considered significant ($\Phi_{ST} = 0.590; p = 0.001$).

Although our data divided the A1/D1 population into two genetic clusters, an additional population structure analysis was performed for lineage A1/D1 West and A1/D1 East separately to reveal putative sub-structuring. For the A1/D1 West population, the ad-hoc statistic $\Delta K$ was maximized for 2 genetic clusters in the population ($K = 2$), indicating that 2 is the most likely number of genetic clusters within the A1/D1 West cluster (Fig. 1; Supporting Information Fig. S4). Here, the Belgian and Dutch isolates formed one genetic cluster that segregated from the French/Japanese isolates. The UK and USA populations both resided in A1/D1 West clusters. In addition, AMOVA analysis also confirmed this sub-division within A1/D1 West as almost 34% of the genotypic variability within A1/D1 West can be explained by this dichotomy ($\Phi_{ST} = 0.338; p = 0.001$). Further genetic clustering of the A1/D1 West population ($K > 2$) did not reveal any more subdivisions in A1/D1 West as isolates were then assigned to more than one genetic cluster. Furthermore, no further genetic clusters were present in lineage A1/D1 East as isolates were more or less equally subdivided between two clusters when $K = 2$. In conclusion, the V. longisporum lineage A1/D1 population contained three genetic clusters with no apparent intermixing between clusters, although intermixing between genetic clusters was enabled in the Structure analysis. The lack of intermixing between the genetic clusters within lineage A1/D1 indicates an exclusively clonal reproduction (Fig. 1). The standardized index of association $R^2$ was calculated to investigate linkage disequilibrium between loci of V. longisporum lineage A1/D1. $R^2$ was significantly different from 0 ($R^2 = 0.3081$, $p < 1.00 \times 10^{-5}$) indicating that V. longisporum is not out-crossing.
Population diversity and genealogy

The *V. longisporum* A1/D1 collection contained 40 different MLGs derived from 79 isolates with a complete genotype, of which 31 were unique (Table 3). The diversity between the UK isolates was higher than between the German and Swedish ones as Nei’s corrected gene diversity ($H_s$) was 0.163, 0.110 and 0.085, respectively (Table 3). In agreement with this, the diversity of the whole A1/D1 West ($H_s = 0.176$) cluster was higher than A1/D1 East ($H_s = 0.109$). The difference in diversity between A1/D1 West and A1/D1 East is also clearly depicted in the genealogical network of the isolates; excluding locus VD12 to reduce the total amount of MLGs to 31 MLGs (Fig. 2). The A1/D1 West and A1/D1 East populations were segregated from each other by a minimal of 10 mutations between MLG 3 and MLG 17. The A1/D1 East network was centred on the modal MLG 6 that represents more than half of the isolates ($n = 22$) from four different countries (France, Germany, Sweden and Latvia). In contrast, A1/D1 West had a less centralized population network with MLGs 15, 23 and 26 being the most represented with 8, 9 and 7 individuals, respectively. MLG 15 contains exclusively UK isolates, whereas MLG23 and MLG26 have representatives from multiple countries (Belgium, Netherlands, UK and USA).

Population origin

The origin of the dichotomy within lineage A1/D1 (Fig. 1) can either point to two independent hybridization events, giving rise to the two sub-populations within A1/D1, or to a single hybridization event followed by evolutionary diversification, leading to the emergence of two sub-populations. To gather additional evidence supporting either hypothesis,
we selected one *V. longisporum* isolate of A1/D1 East (VL20) and one of A1/D1 West (VLB2) for single-molecule real-time (SMRT) sequencing using the PacBio RSII platform, which has been previously demonstrated to deliver high quality genome assemblies of *V. dahliae* and *Verticillium tricorpus* (Faino et al., 2015; Seidl et al., 2015). We generated 466,673 and 457,986 filtered subreads (~67× coverage) for *V. longisporum* strain VL20 and VLB2 that were assembled into 74 and 83 contigs, respectively (Supporting Information Table S2). Subsequent manual curation yielded a final assembly of 44 and 45 contigs with a total assembled genome length of 72.3 and 72.9 Mb for *V. longisporum* VL20 and VLB2, respectively (Supporting Information Table S2), which represents approximately twice the size of the recently assembled complete, telomere-to-telomere assemblies of *V. dahliae* strains JR2 (36.2 Mb) and VdLs17 (36.0 Mb) (Faino et al., 2015).

To place the sequenced *V. longisporum* strains into the context of other *Verticillium* species, we extracted the alleles of four protein-coding genes, namely *actin* (*ACT*), *elongation factor 1-alpha* (*EF*), *glyceraldehyde-3-phosphate dehydrogenase* (*GPD*) and *tryptophan synthase* (*TS*) (Inderbitzin et al., 2011b), from the two genome assemblies and performed maximum likelihood phylogenetic analyses (Supporting Information Fig. S6), confirming that both *V. longisporum* isolates belong to the A1/D1 lineage.

In order to obtain further evidence for their evolutionary origin and to fully utilize the genome assemblies of the two *V. longisporum* isolates, we performed whole-genome comparisons between *V. longisporum* VL20 and VLB2 (Fig. 3A). Whole-genome alignments between *V. longisporum* VL20 and VLB2 revealed large-scale synteny between both genomes, where a single genomic region in one of the genome aligns to two regions in the other genome. Moreover, as expected from an interspecific hybrid, the alignments of one of these regions generally displays >99% identity, while the identity of the second region is lower, ranging from 90 to 95% (Fig. 3B). Notably, comparisons between *V. longisporum* strains VL20 and VLB2 only identified 140 kb and 450 kb, respectively, of genomic material that is absent in the other strain (Fig. 3B). Furthermore, ~1000 SNPs and ~3800 indels between VL20 and VLB2 were revealed. Thus, the two *V. longisporum* strains are genetically highly similar and do not display marked differences between their individual lineage-specific (LS) regions; a likely scenario if the two *V. longisporum* strains emerged from the same hybridization event.

To obtain further evidence the evolutionary origin of the two *V. longisporum* strains, we extracted the individual A1 and D1 sub-genomes based on their sequence identity to *V. dahliae* strain JR2, as parent D1 is presumed to belong to *V. dahliae* (Supporting Information Fig. S6) (Inderbitzin...
Moreover, genome comparisons between *V. dahliae* and *V. longisporum* indicated that 95% sequence identity allows discriminating A from D sub-genome. The extracted A1 and D1 sub-genomes of *V. longisporum* strain VL20 comprised 32.8 Mb and 34.2 Mb, respectively, while 5.2 Mb could not be assigned. Similarly, the A1 and D1 sub-genomes of *V. longisporum* strain VLB2 comprised 32.5 Mb and 34.7 Mb, respectively, while 5.7 Mb could not be assigned. As expected, within and between the *V. longisporum* strains the average identity between A1 and D1 sub-genomes was $\geq 93\%$ (Fig. 3C). Notably, however, the average identity between the A1 sub-genomes as well as between the D1 sub-genomes of both *V. longisporum* strains was $\geq 99\%$ (Fig. 3C), suggesting that the A1 and D1 parental genomes that give rise to *V. longisporum* strain VL20 and VLB2 were identical and that, consequently, A1/D1 East and A1/D1 West are derived from a single hybridization event (Fig. 4).

**Discussion**

Population genetic structures reveal information about the evolutionary history of organisms. Genome hybridization can be a major driver for organismal adaptation and has allowed *V. dahliae* that infects brassicaceous species relatively infrequently, to become pathogenic on these hosts as *V. longisporum* (Depottet et al., 2016a). Based on phylogenetic analyses, *V. longisporum* has been subdivided into three lineages: A1/D1, A1/D2 and A1/D3, each representing a separate hybridization event between two *Verticillium* species (Inderbitzin et al., 2011b). The Verticillium stem striping pathogen has been emerging as a disease on oilseed rape and is emerging in hitherto unaffected production regions (Gladders et al., 2011; CFIA, 2015). This disease is predominantly caused by the *V. longisporum* lineage A1/D1, which is the most virulent lineage on this crop (Novakazi et al., 2015). In this study, the diversity within the A1/D1 lineage was assessed and population structures within a collection from a diverse geographic origin were elucidated. Isolates from nine different countries (Supporting Information Table S1) were genotyped with newly and previously characterized polymorphic SSR loci (Tables 1 and 2). Model-based clustering revealed a hitherto undiscovered dichotomous structuring within *V. longisporum* lineage A1/D1 (Fig. 1 and Supporting Information Fig. S1). Interestingly, the two A1/D1 sub-populations were geographically correlated and were accordingly labelled ‘A1/D1 West’ and ‘A1/D1 East’ based on their relative European location. Geographic population structuring typically indicates adaptation to local climate or to local hosts. Alternatively, the two genetic clusters may also represent two separate hybridization events. In order to test the latter hypothesis, whole-genome comparisons between two representative genomes of the A1/D1 West and East population were performed. All our evidence points towards a single origin of the two A1/D1 populations (Fig. 4). First, the genome sizes of strains VL20 and VLB2 are highly similar, with 72.3 Mb and 72.9 Mb, respectively.
(Supporting Information Table S2). Then, the two genomes carried only a small proportion of lineage specific sequence: 140 kb and 450 kb for VL20 and VLB2, respectively. Recently, genome-comparisons between multiple V. dahliae strains have revealed the presence of extensive (2.5–4.5 Mb) LS regions that are shared by only a subset of the V. dahliae isolates and that enriched for in planta induced genes that contribute to fungal virulence (de Jonge et al., 2013; Faino et al., 2016). For example, direct comparisons between the completely assembled genomes of V. dahliae strains JR2 and VdLS17 (Faino et al., 2015), two strains that have been shown to be extremely closely related with 99.98% sequence identity (de Jonge et al., 2013), identified four large regions comprising ~2 Mb that display frequent presence/absence polymorphisms (Faino et al., 2016), of which in total ~550 kb and ~620 kb in V. dahliae strain JR2 and VdLS17 do not align to the other strain, respectively. Additionally, genome-wide comparison between V. dahliae strains JR2 and VdLS17 revealed ~4700 SNPs and ~10 000 indels, while the V. longisporum VL20 and VLB2 strains contain genomes that were relatively low in divergence as only ~1000 SNPs and ~3800 indels were found. Finally, the genetic distance between the A sub-genomes as well as between the D sub-genomes of both V. longisporum strains was > 99% (Fig. 3C), suggesting that the A and D parental genomes were nearly identical. Thus, we conclude that the observed geographic population structuring is a signature of divergent evolution driven by environmental adaptation.

Presently, environmental conditions responsible for the dichotomous A1/D1 population structure remain elusive. However, strikingly, although all UK A1/D1 isolates were isolated from the same host species, oilseed rape (Supporting Information Table S1), they are more diverse than the German, Swedish and even the whole A1/D1 East population (Table 3). The high diversity of the British population contradicts a recent introduction as genetic bottlenecks can be expected, especially when considering the isolated character of the British island (Nei et al., 1975). Consequently, the A1/D1 West population may have been present in the UK for a longer time without being noticed as Verticillium stem striping pathogen on oilseed rape. Nevertheless, genetic diversity may also be the consequence of multiple introductions. For example, three genetic lineages of the chestnut blight fungus Cryphonec- tria parasitica were found to be separately introduced in France (Dutech et al., 2013).

The A1/D1 West sub-population, partly consisting of UK isolates, is more diverse than its A1/D1 East counterpart (Table 3, Fig. 2). Their shared origin along with this difference in diversity indicates that A1/D1 East is a founder population of the originating A1/D1 West population. The more recent origin of A1/D1 East is also depicted in the genealogical network (Fig. 2). A1/D1 East has a clear modal MLG (MLG6) with all other A1/D1 East MLGs centred on it. In contrast, the A1/D1 West population lacks a dominating genotype. A1/D1 East may have evolved from A1/D1 West in order to adapt to different climate conditions, hence the geographic correlation of the two A1/D1 sub-populations. Obviously, population segregation in organisms with symbiotic relationships can be host driven. Genetic co-structuring between pathogen and host can be found due to their co-evolutionary interaction (Croll and Laine, 2016; Feurtey et al., 2016), as pathogens engage in arm races with hosts for continued symbiosis (Cook et al., 2015). Host-pathogen evolution, also with host plants other than oilseed rape and thus even differences in host range compositions, may explain the discrepancy in population diversity and the Verticillium stem striping emergences of the two A1/D1 sub-populations. Interestingly, Verticillium stem striping has been conspicuous in Sweden and Germany since the 1960s, countries with an exclusive A1/D1 East presence (Fig. 1) (Kroeker, 1970; Daebler et al., 1988). Although we cannot exclude that the A1/D1 West lineage was present as an unnoticed oilseed rape pathogen for a long time, this lineage recently emerged as a conspicuous pathogen on oilseed rape in the UK (Fig. 1) (Gladders et al., 2011). Oilseed rape must have become susceptible, or at least more sensitive, to A1/D1 West for reasons that presently remain unclear. Considering the large degree of diversity within the A1D1 West population, rather than particular adaptations in the pathogen population, alterations in environmental conditions (e.g. global warming, Siebold and Tiedemann, 2012) and oilseed rape cultivars are more likely explanations for the sudden rise of the genomic diverse UK Verticillium stem striping population. Oilseed rape is a relatively new crop in the UK that was virtually unknown in the 1970s but is currently one of the main arable crops with approximate production area of 700 000 hectares (Wood et al., 2013). Intensive breeding efforts have been made to create a broad diversity of oilseed rape oil, such as the modern double low (oo, canola) type (Wittkop et al., 2009). The novelty of Verticillium stem striping makes that cultivar resistance for this disease was not selected for, hence the possibility that more susceptible oilseed rape varieties have been commercialized over time.

V. longisporum is an interspecific hybrid between two separate Verticillium species (Inderbitzin et al., 2011b; Depotter et al., 2016a). Interspecific hybrids are regularly found to be impaired in their sexual reproduction (Greig, 2009; Bertier et al., 2013), although this should be of little significance for V. longisporum as a sexual stage has not been described for any of the Verticillium species (Short et al., 2014). However, mating types, meiosis-specific genes and genomic recombination between clonal lineages have been observed for V. dahliae (Milgroom et al., 2014; Short et al., 2014). This suggests that V. dahliae
may have cryptic or ancestral sexual reproduction. In contrast to population structure studies with *V. dahliae* (Atallah et al., 2010; 2012), no apparent intermixing between genetic clusters was observed for *V. longisporum*. Moreover, the $f^2$ was also significantly different from 0 for the *V. longisporum* lineage A1/D1 ($f^2 = 0.3081, p < 1.00 \times 10^{-5}$), which implies that no linkage equilibrium is present. These data indicate that *V. longisporum* reproduces exclusively in a clonal fashion and has never experienced sexual reproduction. Mechanisms different from meiotic recombination must contribute to genetic diversity in order to achieve evolutionary adaptation (Seidl and Thomma, 2014). Similar as proposed for several other filamentous pathogens, *V. dahliae* has a two-speed genome with particular genomic (LS) regions that are considerably more dynamic than the core genome (Faino et al., 2016). These rapidly evolving LS regions are enriched in active transposable elements that shape the genome in an active and passive fashion (de Jonge et al., 2013; Faino et al., 2016). Moreover, also horizontal gene transfer contributes to genome evolution of *Verticillium dahliae*, as Ave1, an effector gene crucial for aggressiveness, originated by horizontal gene transfer from plants (de Jonge et al., 2012).

Increasing evolutionary ecology knowledge of a pathogen can play a pivotal role in disease management in order to protect ecosystems (Williams, 2010). Population genetic studies are therefore often used to identify origins of newly emerging pathogens and their adaptation pathways (Dutech et al., 2012; Gross et al., 2014). However, not all emerging diseases are preceded by a clearly identified recent introduction. Our study exemplifies that new diseases can emerge from a latent, previously established, microbial population. Thus, factors like weather and farming techniques can also spur disease emergences, especially for fungal pathogens (Anderson et al., 2004). Besides, the two *V. longisporum* populations are generally still bound to specific geographic locations in Europe. The driving forces causing the dichotomous population structure remain elusive, but conceivable differences in pathogenic traits may have contributed to the West/East segregation. A cautious approach is therefore appropriate to prevent further spread of the two populations, as de expansion of the population into new geographic regions may have unpredicted outcomes.

**Experimental procedures**

**Isolate collection, DNA extraction and lineage characterization**

In total 88 isolates from nine different countries were used in this study (Supporting Information Table S1). Mycelium was harvested from two-week-old potato dextrose broth cultures and DNA was extracted according to DNA extraction protocol A from Ribeiro and Lovato (2007). The lineage and mating type to which the respective isolates belong was determined previously for several isolates (Supporting Information Table S1; Inderbitzin et al., 2011b, 2013). The hitherto uncharacterized isolates were screened for the presence of a marker that is specific for lineage A1/D1. To this end, the primer pair D1l/ A1FD1r was used for PCR to amplify a fragment from the GPD locus according to Inderbitzin et al. (2013). Amplicons were displayed by gel electrophoresis on a 1% agarose gel. Furthermore, isolates were screened for MAT1–1 and MAT1–2 idiomorphs with the primer pairs All/MAT11r and HMG21l/ MAT21r, respectively, according to Inderbitzin et al. (2011b). Amplicons were displayed by gel electrophoresis on a 1.5% agarose gel.

**SSR loci**

A genome wide screening for polymorphic SSR loci was done with unpublished draft genome sequences from several *V. longisporum* isolates. INDELS between genomes were extracted from the whole-genome alignments using the mummer package (v3.1) (Kurtz et al., 2004). Gene sequence variations were received by the variant call format tool (Danecek et al., 2011). Insertions and deletions between 5 and 20 nucleotides were selected and screened for recurrent patterns that are typical for SSR loci. Primers were developed for 61 putative polymorphic SSR loci with the Primer3 software (Untergasser et al., 2012). Additional polymorphic SSR markers for lineage A1/D1 were used in this study. VD1, VD8 and VD12 from Atallah et al. (2010) were originally designed for *V. dahliae* and were found to be polymorphic between *V. longisporum* isolates. In addition, VDA783, VDA787 and VDA823 from Barbara et al. (2005), designed for *V. longisporum*, were used. SSR loci were labelled and amplified with an M13 fluorescent tag according to Schuelke (2000). The PCRs consisted of a 2 min initial denaturation step at 95°C, 30 cycles of 35 sec at 95°C, 45 sec at 62°C, and 1 min at 72°C, followed by 8 cycles of 30 sec at 95°C, 45 sec at 53°C and 1 min at 72°C, followed by an extension of 10 min at 72°C. The PCR mix contained 8 pmol of each reverse and M13 tagged universal sequence primer and 2.5 pmol of the forward primer in a final 10 μl reaction volume: 1X GoTaq® Flexi Buffer Mg-free, 2.5 mM MgCl₂, 0.2 mM each dNTP, 100ng template DNA, DNA polymerase, 1.25 u GoTaq® polymerase (Promega, Madison, WI, USA). The labelled PCR products were then combined with Hi-Di formamide and LIZ-500 size standard and resolved on a 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA). The results were processed using GeneMapper v4.0 software (Applied Biosystems, Foster City, CA, USA).

**Population structure**

*V. longisporum* isolates were individually clustered based on polymorphic SSR loci using the software Structure version 2.3 (Pritchard et al., 2000). Data was analysed as for a haploid organism as all markers only gave a single polymorphic signal. The population was tested for containing 1 up to 6 genetic clusters (K). For every cluster, 10 runs were performed with a burn-in period of 500 000 generations and 1 000 000 Markov
Chain Monte Carlo (MCMC) simulations. The admixture model was chosen and the loci were considered independent. The most likely number of genetic clusters in the population was determined with the ad-hoc statistic $\Delta K$ (Evanno et al., 2005) using Structure Harvester (Earl and vonHoldt, 2012). Hereby, the amount of clusters in a population is determined based on the rate of change in the log probability of data between successive $K$ values. Furthermore, the results from Structure were permuted and aligned in the program CLUMPP 1.1.2 (method Greedy, random input order, 1 000 000 repeats) (Jakobsson and Rosenberg, 2007) and visualized with the software Distruct 1.1 (Rosenberg, 2004). The correlation between the clusters and the country of origin of the isolates was determined with the Spearman’s rank correlation coefficient ($\rho$) with Hmisc package in R 3.2.3 (R Core Team, 2015; Harrell and Dupont, 2016). The standardized index of association $f_{is}^2$ is an indication for the recombination rate between organisms. In outcrossing populations, no linkage disequilibrium is present and an $f_{is}^2$ of 0 is expected (Burt et al., 1996). $f_{is}^2$ was computed to test the recombination potential within the *V. longisporum* lineage A1/D1 population using the software LIAN version 3.7 (Haubold and Hudson, 2000). A Monte Carlo simulation of 100 000 iterations was chosen. The genetic clusters identified by Structure were evaluated using an analysis of molecular variance (AMOVA) with the software GenAlEx version 6.502 (Peakall and Smouse, 2000). A Monte Carlo simulation of 100 000 iterations was chosen. The genetic clusters identified by Structure were evaluated using an analysis of molecular variance (AMOVA) with the software GenAlEx version 6.502 (Peakall and Smouse, 2006; 2012). The variance within the genetic clusters was compared with the variance between the genetic clusters with an analogue of Wright’s fixation index ($\theta_{PT}$) (Excoffier et al., 1992). The population diversity was assessed for populations with a minimum of 10 representatives, excluding isolates with missing data. Multi-locus genotype (MLG) diversity and the allelic richness (AR) were determined for each population using software Contrib 1.4, using rarefaction size 5 (Petit et al., 1998). Nei’s (1973) genetic diversity corrected for sample size ($H_{E}$) values were generated in GenoDive (Meirmans and van Tienderen, 2004). Genealogical relationships among the different MLGs haplotypes in the *V. longisporum* population was inferred using the median-joining method (Bandelt et al., 1999), implemented in the software Network 5.0.0.0 (http://www.fluxus-engineering.com). All SSR loci were weighted equally (10) and an epsilon = 0 was chosen. The hypervariable SSR locus VD12 was not included in the analysis to reduce the amount of MLGs, and MLGs with missing data were excluded as well.

**Genome sequencing and assembly of two V. longisporum isolates**

Genomic DNA of *V. longisporum* isolates VL20 and VLB2 was isolated from conidia and mycelium fragments that were harvested from 10-day-old cultures grown in liquid potato dextrose agar according to the protocol described by Seidl et al. (2015). The PacBio libraries for sequencing on the PacBio RSII machine (Pacific Biosciences of California, CA, USA) were constructed using ~20 µg of *V. longisporum* DNA, similarly as described previously by Faino et al., (2015). Briefly, DNA was mechanically sheared, and size selected using the BluePippin preparation system (Sage Science, Beverly, MA, USA) to produce ~20 kb insert size libraries. The sheared DNA and final library were characterized for size distribution using an Agilent Bioanalyzer 2100 (Agilent Technology, Inc., Santa Clara, CA, USA). The PacBio libraries were sequenced on six SMRT cells per *V. longisporum* isolate at KeyGene N.V. (Wageningen, the Netherlands) using the PacBio RS II instrument. Sequencing was performed using the P6-C4 polymerase-Chemistry combination and a > 4 h movie time and stage start. Post filtering, a total of 344 906 (N50 ~20 kb) and 358 083 (N50 ~19 kb) polymerase reads were obtained for *V. longisporum* isolates VLB2 and VL20, respectively. Filtered sub-reads for VL20 and VLB2 (457 986; N50 ~14 kb and 466 673; N50 ~13.7 kb, respectively) were assembled using the HGAP v3 protocol (Chin et al., 2013). Subsequently, the HGAP3 assemblies underwent additional polishing using Quiver (Chin et al., 2013). The *de novo* assemblies were further upgraded using FinisherSC, and the upgraded assemblies were polished with Quiver (Lam et al., 2015). Lastly, contigs that displayed very low or exceptionally high PacBio read coverage, as well as the contig representing the mitochondrial DNA, were removed from the final assemblies. *De novo* repetitive elements in the genomes of *V. longisporum* were identified using RepeatModeler (v1.0.8), and repetitive elements were subsequently masked using RepeatMasker (v4.0.6; sensitive mode).

**Genome comparisons between V. longisporum and V. dahliae strains**

To place the newly sequenced *V. longisporum* isolates in context of 74 previously analysed *Verticillium* spp., we identified the alleles (A or D) of four previously used protein-coding genes actin (ACT), elongation factor 1-alpha (EF), glyceraldehyde-3-phosphate dehydrogenase (GPD) and tryptophan synthase (TS) in the genome assemblies of VL20 and VLB2 using blastn searches (Inderbitzin et al., 2011a). Sequences were extracted from the genome assemblies and aligned to the four genes in the 74 *Verticillium* isolates using mafft (LINSI: v7.271) (Katoh and Standley, 2013). A phylogenetic tree was reconstructed using PhyML using the GTR nucleotide substitution model and four discrete gamma categories (Guindon and Gascuel, 2003). The robustness of the phylogeny was assessed using 500 bootstrap replicates.

Whole-genome comparisons between *V. longisporum* strains VLB2 and VL20 and between *V. dahliae* strains JR2 and VdLS17 (Faino et al., 2015) were performed with nuclmer (maxmatch), which is part of the mummer package (v3.1) (Kurtz et al., 2004). Small polymorphisms (SNPs and INDELs) between genomes were extracted from the whole-genome alignments using show-snps (excluding ambiguous alignments), which is part of the mummer package (v3.1). Lineage-specific regions per individual *Verticillium* strain were determined by extracting nuclmer alignments and subsequently identifying genomic regions that lack alignments with the other isolate (bedtools genomcvcov) (Quinlan and Hall, 2010).

To tentatively assign the individual sub-genomes, the repeat-masked genome of *V. dahliae* strain JR2 was compared to the repeat-masked *V. longisporum* genomes using nuclmer (maxmatch), of which only 1-to-1 alignments longer than 5 kb were retained. Genomic regions were assigned to parental sub-genomes based on the average identity of
consecutive alignments (defined by location and/or strand), where regions with average identity >95% were assigned to D and < 95% identity to A, respectively. The pairwise identity between A and D parents within and between V. longisporum strains was calculated using nucmer (mum), with dividing the respective query sequences into non-overlapping windows of 500 bp.

**Sequence data**

SSR data have been deposited as GenBank accessions KY828946-KY828954. The Verticillium longisporum genome sequence data have been deposited as SRR5430591–6 (VLB2) and SRR5435055–60 (VL20).

**Acknowledgements**

The authors would like to thank the Marie Curie Actions programme of the European Commission that financially supports the research investigating the threat of V. longisporum to UK oilseed rape production. Work in the laboratories of B.P.H.J.T. and M.F.S is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands Organization of Scientific Research (NWO). We greatly appreciate all donors of V. longisporum isolates: the Dixielus, Høfte, INRA-CIRAD, Norddeutsche Pflanzenzucht, Subbarao and Terres Inovia laboratories. The authors declare no conflict of interest.

**References**


**Divergent evolution within Verticillium longisporum**


© 2017 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*, 00, 00–00


Divergent evolution within *Verticillium longisporum* 13


Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

Table S1. *Verticillium longisporum* isolates used in this study.

Table S2. *Verticillium longisporum* strain VL20 and VLB2 genome assemblies.

Fig. S1. Genetic clusters within the *Verticillium longisporum* population using the 9 newly designed polymorphic simple sequence repeat (SSR) markers.

Fig. S2. Lineage characterization of *Verticillium longisporum* isolates.

Fig. S3. Mating type characterization of *Verticillium longisporum* isolates.

Fig. S4. Genetic clusters within the *Verticillium longisporum* A1/D1 population using 13 polymorphic simple sequence repeat (SSR) markers.

Fig. S5. Genetic clusters within the *Verticillium longisporum* A1/D1 West population using 11 polymorphic simple sequence repeat (SSR) markers.

Fig. S6. Phylogenetic relationships between *Verticillium longisporum* A1/D1 West and East parents with the other *Verticillium* species.