Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen

Luigi Faino,1,3 Michael F. Seidl,1,3 Xiaqian Shi-Kunne,1 Marc Pauper,1 Grardy C.M. van den Berg,1 Alexander H.J. Wittenberg,2 and Bart P.H.J. Thomma1

1Laboratory of Phytopathology, Wageningen University, 6708 PB Wageningen, The Netherlands; 2Keygene N.V., 6708 PW Wageningen, The Netherlands

Genomic plasticity enables adaptation to changing environments, which is especially relevant for pathogens that engage in “arms races” with their hosts. In many pathogens, genes mediating virulence cluster in highly variable, transposon-rich, physically distinct genomic compartments. However, understanding of the evolution of these compartments, and the role of transposons therein, remains limited. Here, we show that transposons are the major driving force for adaptive genome evolution in the fungal plant pathogen Verticillium dahliae. We show that highly variable lineage-specific (LS) regions evolved by genomic rearrangements that are mediated by erroneous double-strand repair, often utilizing transposons. We furthermore show that recent genetic duplications are enhanced in LS regions, against an older episode of duplication events. Finally, LS regions are enriched in active transposons, which contribute to local genome plasticity. Thus, we provide evidence for genome shaping by transposons, both in an active and passive manner, which impacts the evolution of pathogen virulence.

[Supplemental material is available for this article.]

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

Although sexual reproduction drives genotypic diversity, not all eukaryotes regularly reproduce sexually, including many asexual fungal phyla (McDonald and Linde 2002; Heitman et al. 2007; Flot et al. 2013). However, in such asexual organisms adaptive genome evolution also occurs, mediated by various mechanisms ranging from single-nucleotide polymorphisms to large-scale structural variations that can affect chromosomal shape, organization, and gene content (Seidl and Thomma 2014). Verticillium dahliae is a soil-borne fungal pathogen that infects susceptible hosts through their roots and colonizes the water-conducting xylem vessels, leading to vascular wilt disease (Fradin and Thomma 2006). Despite its presumed asexual nature, V. dahliae is a highly successful pathogen that causes disease on hundreds of plant hosts (Fradin and Thomma 2006; Klosterman et al. 2009; Inderbitzin and Subbarao 2014). Using comparative genomics, we recently identified genomic rearrangements in V. dahliae that lead to extensive chromosomal length polymorphisms (de Jonge et al. 2013). Moreover, the genomes of V. dahliae strains were found to contain highly dynamic, repeat-rich, lineage-specific (LS) regions (Klosterman et al. 2011; de Jonge et al. 2013). Intriguingly, these LS regions are enriched for in planta-induced effector genes that contribute to fungal virulence (de Jonge et al. 2013). Similarly, many filamentous pathogens are considered to have evolved so-called “two-speed” genomes with gene-rich, repeat-poor genomic compartments that contain core genes that mediate general physiology and evolve slowly, whereas plastic, gene-poor and repeat-rich compartments are enriched in effector genes that mediate virulence in interactions with host plants and evolve relatively quickly (Raffaele and Kamoun 2012).

Plastic, fast-evolving genomic compartments in plant and animal pathogen genomes concern particular regions that are either embedded within the core chromosomes or reside on conditionally dispensable chromosomes (Thon et al. 2006; Fedorova et al. 2008; Haas et al. 2009; Ma et al. 2010; Goodwin et al. 2011; Klosterman et al. 2011; Rouxel et al. 2011; de Jonge et al. 2013). Irrespective of the type of organization of the two-speed genome, the fast-evolving compartment is generally enriched for transposable elements (TEs) that are thought to actively promote genomic changes by causing DNA breaks during excision or by acting as a substrate for rearrangement (Seidl and Thomma 2014). Nevertheless, the exact role of TEs in the evolution of effector genes currently remains unknown (Dong et al. 2015). We recently established gapless V. dahliae whole-genome assemblies using a combination of long-read sequencing and optical mapping (Faino et al. 2015; Thomma et al. 2015). Here, we exploit these novel assemblies for...
in-depth investigations into the molecular mechanisms responsible for genomic variability, which is instrumental for adaptive genome evolution in *V. dahliae*.

**Results**

Genomic rearrangements in *Verticillium dahliae* are associated with sequence similarity

Whole-genome alignments between *V. dahliae* strains JR2 and VdLs17 revealed 24 synteny disruptions (Fig. 1A; Supplemental Fig. S1). To further increase their resolution, we aligned long (average ~9 kb) sequencing reads derived from strain VdLs17 (Faino et al. 2015) to the strain JR2 assembly, and reads of which the two sides aligned to two distinct genomic locations were used to determine breakpoints at high-resolution (Supplemental Fig. S2). After manual refinement, this procedure yielded 19 large-scale chromosomal rearrangements; 13 inter-chromosomal translocations, and six intra-chromosomal inversions (Fig. 1; Table 1). Although seven rearrangements were confined to regions smaller than 100 bp, the remaining 12 concern larger genomic regions that could not be further refined due to the presence of strain-specific or repeat-rich regions (Fig. 1A).

We subsequently assessed the occurrence of the 13 inter-chromosomal rearrangements in nine other *V. dahliae* strains by querying paired-end reads derived from these genomes for pairs of discordantly mapped reads (i.e., both reads fail to map at the expected distance or location) when mapped onto the strain JR2 assembly, revealing distinct rearrangement patterns (Supplemental Fig. S3). Although some synteny breakpoints identified in *V. dahliae* strain JR2 are either specific to *V. dahliae* strain VdLs17 (Chr 1: 2,843,014–2,843,020) (Supplemental Fig. S4A) or common to all other *V. dahliae* strains (Chr 1: 8,013,865–8,013,868) (Supplemental Fig. S4B), some are observed in only a subset of *V. dahliae* strains (Chr 3: 3,130,776–3,130,781) (Supplemental Fig. S4C).

*V. dahliae* strains JR2 and VdLs17 are highly similar with only 8622 SNPs difference (0.024% nucleotide divergence). When focusing on the regions (10 kb) upstream of and downstream from the 19 identified synteny breakpoints, we observe ~200 SNPs (~0.05% divergence), indicating a moderate increase in proximity of the breakpoints. Similarly, when considering all *V. dahliae* strains, we observe a genome-wide average of 4.1 SNPs/kb; whereas around breakpoints (10 kb upstream and downstream), 6.3 SNPs/kb are found.

Genomic rearrangements are generally caused by double-strand DNA breakages followed by unfaithful repair by mechanisms that utilize homologous sequences to repair such breaks (Hedges and Deininger 2007; Krejci et al. 2012; Seidl and Thomma 2014). Repetitive genomic elements, such as transposable elements (TEs), may give rise to genomic rearrangements by providing an ectopic substrate that interferes with faithful repair of the original break. Notably, of 19 chromosomal rearrangements identified in *V. dahliae* strain JR2, 15 colocalize with a repetitive element (Table 1; Supplemental Table S1). Of the 13 inter-chromosomal rearrangements, 12 could be reconstructed in detail (Fig. 1A; Table 1; Supplemental Figs. S5, S6). Eight of them occur over highly similar TEs in both strains, although they belong to different TE families (Fig. 1B; Table 1; Supplemental Figs S5, S6). At two of the breakpoints, a TE occurred in strain VdLs17 that was absent at those breakpoints in the JR2 strain. For the final two breakpoints, no association to a TE was found in either strain, but extended sequence similarity surrounding the rearrangement site was identified (Fig. 1B; Supplemental Fig. S6). Therefore, we conclude that not necessarily TEs or their activity, but rather stretches of sequence similarity, are associated with ectopic chromosomal
Table 1. Genomic location of synteny breakpoints identified in *Verticillium dahliae* strain JR2 and the closest flanking genomic element

| Chr  | Start       | End        | Interval (bp) | Type | Start       | End        | Strand | Genomic element | Distance to breakpoint (bp) | Index | Rearrangement number |
|------|-------------|------------|---------------|------|-------------|------------|--------|-----------------|-----------------------------|-------|---------------------|
| 1    | 2,843,014   | 2,843,020  | 6             | TRA  | 2,840,162   | 2,843,030  | +      | VDAG_JR2_Chr1g08970 | 0               | a     | 1                   |
| 1    | 7,140,631   | 7,162,273  | 21,643        | INV  | 7,162,323   | 7,162,992  | +      | VdLTRE8        | 0               | 0     |                     |
| 1    | 7,227,694   | 7,243,623  | 15,930        | INV  | 7,242,070   | 7,248,997  | +      | VdTRE12        | 0               | 0     |                     |
| 1    | 8,013,865   | 8,013,868  | 3             | TRA  | 8,013,244   | 8,014,018  | -      | VDAG_JR2_Chr1g25440 | 0               | b     | 2                   |
| 2    | 2,876,424   | 2,897,525  | 21,102        | TRA  | 2,896,858   | 2,898,045  | +      | VdUNK65        | 0               | c     | 3                   |
| 2    | 3,025,499   | 3,031,900  | 6,401         | TRA  | 3,030,927   | 3,030,993  | -      | VdUNK35        | 0               | d     | 4                   |
| 2    | 3,120,447   | 3,122,897  | 2451          | TRA  | 3,122,730   | 3,124,308  | +      | VdUNK65        | 0               | e     | 5                   |
| 2    | 3,334,994   | 3,620,790  | 285,796       | TRA  | 3,620,790   | 3,629,961  | -      | VdTRE10        | 0               | f     | 6                   |
| 2    | 4,225,478   | 4,225,575  | 97            | TRA  | 4,225,210   | 4,226,423  | -      | VDAG_JR2_Chr2g13160 | 0               | b     | 7                   |
| 3    | 149,893     | 166,205    | 16,313        | INV  | 164,460     | 166,293   | +      | VdUNK19        | 0               | 0     |                     |
| 3    | 226,642     | 226,647    | 5             | INV  | 225,927     | 226,658   | +      | VdTRE8        | 0               | 0     |                     |
| 3    | 3,130,776   | 3,130,781  | 5             | TRA  | 3,129,525   | 3,130,711  | -      | VDAG_JR2_Chr3g10000 | 64              | a     | 8                   |
| 4    | 1,464,458   | 1,480,146  | 15,689        | INV  | 1,479,722   | 1,479,771  | +      | 5SrRNA-1_BG    | 0               | 0     | 9                   |
| 5    | 2,171,284   | 2,202,070  | 30,787        | TRA  | 2,195,136   | 2,202,070  | +      | VdTRE2_1      | 0               | 0     | 10                  |
| 5    | 2,367,806   | 2,390,749  | 22,944        | INV  | 2,390,710   | 2,390,856  | +      | VdUNK100       | 0               | 0     |                     |
| 5    | 3,423,150   | 3,423,640  | 491           | TRA  | 3,422,285   | 3,423,644  | +      | VdTRE12        | 0               | d     | 10                  |
| 5    | 3,642,999   | 3,644,451  | 1,453         | TRA  | 3,644,182   | 3,650,054  | +      | VdTRE1_1      | 0               | f     | 11                  |
| 5    | 3,686,406   | 3,686,412  | 6             | TRA  | 3,685,349   | 3,686,592  | -      | VdUNK11        | 0               | c     | 12                  |
| 6    | 3,218,241   | 3,218,279  | 39            | TRA  | 3,212,527   | 3,220,016  | +      | VdTRE10        | 0               | e     | 13                  |

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

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

*c* The closest genomic elements comprise genes and repeats (see also Supplemental Table S1).

*d* The letter label (Index) indicates the panel in Supplemental Figure S5 in which the particular rearrangement is graphically displayed.

*e* The number (Rearrangement number) links the syntenic breaks to the rearrangements represented in Figure 1A.
rearrangements in *V. dahliae*, likely mediating unfaithful homology-based DNA repair. Since TEs are more abundant compared to other (types of) sequences, these are more likely to become substrates for double-strand repair pathways.

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

Whole-genome alignments between *V. dahliae* strains JR2 and VdLs17 revealed four large (>10 kb) repeat-rich genomic regions lacking synteny between the strains (Supplemental Table S2). Mapping of reads from nine additional strains (de Jonge et al. 2012) onto the JR2 and VdLs17 assemblies revealed absence of read coverage primarily at their respective LS regions (Supplemental Fig. S7). Notably, several inter-chromosomal rearrangements colocalize with these LS regions (Fig. 1A), suggesting that genomic rearrangements contributed to the formation of LS regions.

Although duplicated genes have previously been observed in LS regions of *V. dahliae* (Klosterman et al. 2011), the extent of such duplications and their role in the evolution of LS regions remains unknown. To determine the extent of segmental duplications in LS regions, we used two approaches. First, we performed whole-genome nucleotide alignments of LS regions, we used two approaches. First, we performed whole-genome alignments between *V. dahliae* and VdLs17. The first duplication period that occurred before the last common ancestor, where-as the most recent split was the divergence of *Verticillium* alfalfae. Genes in the LS regions have previously been shown to be particularly relevant for pathogen virulence in

![Figure 2](www.genome.org)
**V. dahliae** (de Jonge et al. 2013), suggesting that recent gene duplications are contributing to the evolution of virulence.

The recent duplications that affected LS regions have generated raw genetic material that can be subjected to subsequent rapid evolutionary diversification, leading to novel or altered gene functionality, but can also be subject to differential loss of one of the duplicated gene copies (Fig. 3; Supplemental Fig. S9). In general, LS regions in **V. dahliae** strain JR2 display considerable gene loss, because for about 100 of the approximately 400 genes located in LS regions, no ortholog could be detected in **V. dahliae** strain VdLs17. Thus, differential gene loss significantly contributed to the diversification of LS regions. To determine if LS regions display signs of increased gene diversification and selection pressure acting on protein-coding genes, we used the ratio of nonsynonymous substitutions per nonsynonymous site (K_a) as well as the rate of increased K_a to K_s values calculated between orthologous gene pairs of **V. dahliae** strain JR2 and the closest related **Verticillium** species, **V. alfalfae**, as a proxy. Genes located at LS regions display moderately increased K_a values, as well as K_a/K_s values, when compared to genes residing in the core genome (median of 0.07 compared to 0.03, and 0.37 compared to 0.18, for LS and core genes, respectively). Although these indicate accelerated sequence divergence of genes located within LS regions, the moderate differences also corroborate previous results (de Jonge et al. 2013; Seidl et al. 2015), suggesting that sequence divergence only plays a minor role in **V. dahliae** genome evolution.

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

As shown above, highly dynamic LS regions are characterized by frequent gene duplications and differential gene loss. Moreover, effector genes located in LS regions play decisive roles in pathogen-host interactions (de Jonge et al. 2013). For example, Ave1 is an important LS effector that determines **V. dahliae** virulence on various host plants (de Jonge et al. 2012). As expected in a coevolutionary “arms race” (Thomma et al. 2011; Cook et al. 2015), host recognition evolved as host plants that carry the Ve1 immune receptor recognize this effector (Fradin et al. 2009; de Jonge et al. 2012). Whereas race 1 strains of **V. dahliae** are contained by the Ve1 immune receptor, race 2 strains evade recognition due to loss of Ave1 and are able to infect Ve1 host plants.

In **V. dahliae** strain JR2, Ave1 is embedded in a gene-sparsely and repeat-rich LS region on Chromosome 5 (approximately 550,000–1,050,000) (Fig. 4). Notably, the average number of single nucleotide polymorphisms (SNPs) inferred from three race 1 and eight race 2 strains of **V. dahliae** is significantly reduced in the area surrounding the Ave1 locus (between 680,000 and 720,000) when compared with the surrounding genomic regions (Fig. 4B), but also compared to genome-wide SNP levels in LS and core regions (Supplemental Fig. S10), which may be explained by the relatively recent acquisition of this region through horizontal transfer (de Jonge et al. 2012).

**V. dahliae** race 1 strains do not occur as a monophyletic clade in the **V. dahliae** population (Supplemental Fig. S3A; de Jonge et al. 2013). To assess if Ave1 was gained or lost multiple times, we studied the genomic region surrounding the Ave1 locus in the LS region. By mapping paired-end reads derived from genomic sequencing of various **V. dahliae** strains onto the genome assembly of **V. dahliae** strain JR2, we observed clear differences in coverage levels between **V. dahliae** race 1 and race 2 strains that carry or lack the Ave1 gene, respectively (Fig. 4C; Supplemental Fig. S11). Although **V. dahliae** race 1 strains, including JR2, displayed an even level of read coverage over the Ave1 locus, no read coverage over the Ave1 gene was found in race 2 strains. Intriguingly, read coverage surrounding the Ave1 gene revealed that race 2 strains can be divided into three groups, depending on the exact location of the read coverage drop (Fig. 4C; Supplemental Fig. S11B). Whereas one group of isolates does not display any read coverage over a region of ∼40 kb flanking the Ave1 gene (Supplemental Fig. S11B, colored lines from ∼680 to ∼720 kb), two groups display distinct regions in which the read coverage around the Ave1 locus drops (Supplemental Fig. S11B, red lines around 668 kb and green lines around 672 kb). In conclusion, the Ave1 locus is situated in a highly dynamic and repeat-rich region (Fig. 4C), and the most parsimonious evolutionary scenario is that the Ave1 locus was
horizontally acquired once, followed by multiple losses in independent lineages that encountered host plants that carried Ve1 or functional homologs of this immune receptor (Thomma et al. 2011; de Jonge et al. 2012; Zhang et al. 2014; Song et al. 2016).

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

Although the activity of TEs is not associated with the formation of extensive genome rearrangements, LS regions are highly enriched for TEs, and their presence and potential activity may contribute to accelerated evolution of these genomic regions. In *V. dahliae*, the most abundant class of TEs are retrotransposons that transpose within the genome using an RNA intermediate (Supplemental Table S1; Wicker et al. 2007; Faino et al. 2015). We assessed TE dynamics by querying the transcriptional activity of TEs using in vitro RNA-seq data derived from *V. dahliae* strain JR2 (de Jonge et al. 2012). Notably, the majority of TEs in *V. dahliae* are not transcribed and thus likely not active (Supplemental Table S1), whereas transcribed and therefore likely active TEs are found in LS regions—171 of 280 TEs in LS regions (61%) are expressed (log_{10}(RPKM + 1) > 0) (Supplemental Fig. S12).

To further assess if and how TEs influence the evolution of LS regions, we explored TE dynamics in the genome of *V. dahliae* strain JR2. Each copy of a TE in the genome is derived from an active ancestor that, once transposed and integrated into the genome, accumulates mutations that over evolutionary time will render the TE inactive. The relative age of individual TEs can thus be estimated based on sequence divergence from a consensus sequence that can be derived from present-day copies of any given TE. Using the Jukes-Cantor distance (Jukes and Cantor 1969), which corrects the divergence between TEs and their consensus sequence for multiple substitutions, we estimated the divergence times for TEs in the *V. dahliae* strain JR2 genome (Fig. 5). This analysis showed that TEs primarily transposed and expanded in two distinct periods (Fig. 5A). Notably, a considerable amount of “younger” TEs, i.e., with small Jukes-Cantor distance (Jukes and Cantor 1969) to their consensus sequence, localizes in LS regions, whereas the majority of “older” TEs resides in the core genome (Fig. 5A). In line with this observation, 90 of 280 TEs in LS regions (i.e., 32%) display <0.025 nucleotide divergence from the consensus sequence, whereas 151 of 1101 TEs in the core (i.e., 14%) display such low levels of divergence. Similarly, genome-wide, 202 of the 390 transcriptionally active TEs (i.e., 52%) display <0.025 divergence, whereas only 39 of 991 transcriptionally silent TEs (i.e., 4%) display such low levels of divergence (Fig. 5B). Additionally, the relative age of individual TEs belonging to the LTR class was estimated based on sequence divergence between the distal LTR sequence motifs using the Jukes-Cantor distance (Fiston-Lavie et al. 2012). Despite the relatively low amount of complete LTR-type TEs found in the genome of *V. dahliae* strain JR2, this analysis showed a similar age distribution (Supplemental Fig. S13). Next, we attempted to determine the relative period in which the majority of TEs in the genome of *V. dahliae* strain JR2 transposed. To this end, we derived Jukes-Cantor distributions for orthologous genes of *V. dahliae* strain JR2 and individual closely related fungi as a proxy of divergence between these species (Fig. 5C). These distributions display the same pattern of species divergence when derived from phylogenetic analyses (Supplemental Fig. S14) as well as from the K distances between orthologous gene pairs (Fig. 2C). By comparing the Jukes-Cantor distributions derived from TEs and from orthologous genes, we revealed that older TEs transposed before the separation of *V. dahliae* and *V. alfalfa*, whereas the younger TEs transposed after *V. dahliae* speciation (Fig. 5C). Notably, younger TEs tend to be transcriptionally active, and older, more diverged TEs tend to be transcriptionally silent (Fig. 5B). Thus, the expansion of younger TEs is recent and primarily concerns the active TEs localized at LS regions (Fig. 5), strongly suggesting that TE transpositions contribute to the genetic plasticity of LS regions.

**Discussion**

It was previously shown that genetic rearrangements occur in fast-evolving LS regions of the genome that are enriched in TEs (Klosterman et al. 2011; de Jonge et al. 2013). We now show that TEs are enriched near rearrangement breakpoints and play a role in homology-based recombinations. We furthermore show that recent genetic duplications are enhanced in LS regions, against the background of an older episode of duplication events for the “core” genome. Finally, we show that LS regions display signs of TE activity that may contribute to plasticity of those regions.

Many plant pathogens contain a so-called “two-speed” genome, where effector genes reside in genomic compartments
Figure 5. Dynamics of transposable elements in the genome of *Verticillium dahliae* strain JR2. (A) The divergence time of transposable elements identified in the genome of *V. dahliae* strain JR2 (Faino et al. 2015) was estimated using the Jukes-Cantor distance calculated between repeat copies and their consensus sequence. The distributions of divergence times between transposable elements located in the core genome (red) and in the LS regions (blue) differ. Estimations of speciation events in the evolutionary history of *V. dahliae* are indicated by triangles based on analyses in C. (B) The distributions of divergence times between expressed/active (log10[RPKM+1] >0) transposable elements (red) and nonexpressed (blue) transposable elements differ. Estimations of speciation events are indicated by triangles. (C) Speciation events are estimated by calculating the Jukes-Cantor distance for orthologous gene pairs based on genes from *V. dahliae* strains JR2 and their respective orthologs in the other genomes. Distributions and median divergence times between 1:1 orthologous pairs, displayed by box plots, were used to estimate relative speciation events.

Intriguingly, although the majority of TEs in *V. dahliae* is transcriptionally silent and thus inactive, transcriptionally active and thus likely actively transposing TEs were observed in LS regions. This observation was corroborated using a very conservative approach to map RNA-seq reads to limit the effect of reads mapping to multiple locations in the genome. Transcription and specific induction of TE activity has been observed previously in *V. dahliae* strain VdLs17 (Klosterman et al. 2011), for which we recently revealed a considerable number of erroneous chromosomal inversions (Faino et al. 2015), we still observed that three of four large-scale LS regions are associated with chromosomal rearrangements. We were not able to associate every LS region with a genomic rearrangement, and we were also not able to exactly reconstitute each genomic rearrangement. However, it can be anticipated that complex rearrangements occurred in which genetic material in close proximity was lost. Furthermore, rearrangements may have occurred over longer evolutionary timescales, and subsequent rearrangement events may have erased “scars” of previous rearrangements. Thus, although not every genomic rearrangement can be associated to an LS region, large-scale genomic alterations appear to be the driving force for LS region formation in *V. dahliae*.

Here, we identified ~2-Mb repeat-rich, lineage-specific (LS) regions between two *V. dahliae* strains that are significantly enriched in TEs and contain all thus far functionally analyzed effector genes, including *Avr1*. Moreover, we also determined a significant number of syntenic breakpoints that are associated with genomic rearrangements to high resolution, most of which we could reconstitute in detail. Although the previously determined association between LS compartments and the occurrence of genomic rearrangements was overestimated due to major errors in the publicly available genome assembly of *V. dahliae* strain VdLs17 (Klosterman et al. 2011), for which we recently revealed a considerable number of erroneous chromosomal inversions (Faino et al. 2015), we still observed that three of four large-scale LS regions are associated with chromosomal rearrangements. We were not able to associate every LS region with a genomic rearrangement, and we were also not able to exactly reconstitute each genomic rearrangement. However, it can be anticipated that complex rearrangements occurred in which genetic material in close proximity was lost. Furthermore, rearrangements may have occurred over longer evolutionary timescales, and subsequent rearrangement events may have erased “scars” of previous rearrangements. Thus, although not every genomic rearrangement can be associated to an LS region, large-scale genomic alterations appear to be the driving force for LS region formation in *V. dahliae*.
et al. 2000; Chuma et al. 2011). Similar processes likely contribute to the frequent recovery of Avr-Pita in M. oryzae strains relieved from selection pressure of Pita, leading to translocation and duplications of this effector gene in dynamic genomic regions (Chuma et al. 2011).

Genome-wide studies in several fungi aiming to study chromatin revealed that TE-rich regions are generally associated with highly condensed chromatin that restricts transcription and TE activity (Lewis et al. 2009; Connolly et al. 2013; Galaiza and Freitag 2014). Therefore, TEs can influence the expression of neighboring genes such as effectors, as they can direct the formation of heterochromatic regions (Lewis et al. 2009). In the saprophytic fungus Neurospora crassa, heterochromatin formation at TEs is directed by remnants of Repeat Induced Point mutations (RIP), a premeiotic process that actively induces point mutations in TEs (Selker 1990; Lewis et al. 2009, 2010). In the pathogenic fungus Leptosphaeria maculans, effectors are located in TE-rich regions that were subjected to extensive RIP mutations and display signatures of nucleotide mutations caused by the RIP process, fostering rapid effector diversification (Fudal et al. 2009; Daverdin et al. 2012). However, as Verticillium spp. are considered asexual, it is not unexpected that effectors in LS regions do not display evidence for RIP mutations. Despite the fact that LS regions do not show overrepresentation of secreted genes, we previously observed that effector genes residing in the LS regions are considerably over-represented in the V. dahliae transcriptome upon plant infection (de Jonge et al. 2013). Potentially, since several TEs in LS regions are transcriptionally active, LS regions may either not yet be targeted by heterochromatin formation or carry different chromatin marks as core genomic compartments. Chromatin-based regulation of effector genes and genes encoding other virulence factors has been observed in several pathogenic fungi (Connolly et al. 2013; Qutob et al. 2013; Chujo and Scott 2014; Soyer et al. 2014, 2015). Thus, further insight into chromatin biology of V. dahliae will be instrumental to significantly enhance our knowledge of the evolution of its two-speed genome and on virulence.

Methods

Genome annotation and dynamics

Gene predictions for the genome assembly of V. dahliae strain JR2 (Faino et al. 2015) were performed using Maker2 software (Holt and Yandell 2011) with multiple independent evidence, including transcriptomic data, to support gene annotation (Klosterman et al. 2011; de Jonge et al. 2012, 2013; Seidl et al. 2015). For details, see Supplemental Methods. Homology between protein-coding genes of 13 fungal species was assessed using OrthoMCL (Li et al. 2003). Orthologous gene pairs between fungal species and paralogous gene pairs within each individual species were extracted from the OrthoMCL gene families. $K_a$ values between gene pairs, as defined by OrthoMCL families, were calculated using the $K_a$-$K_s$ Calculator 2.0 package (Wang et al. 2010).

Repetitive elements were identified as described in Faino et al. (2015). Expression of repetitive elements was assessed based on RNA sequencing data derived from V. dahliae strain JR2 grown in in vitro media (Czapek Dox) (de Jonge et al. 2012). Paired end reads were mapped onto the genome assembly of V. dahliae strain JR2 using TopHat2 (Trapnell et al. 2009), summarized, and reported as reads per kilobase of transcript per million mapped reads (RPKM). To estimate divergence time of transposable elements, each individual copy of a transposable element was aligned to the consensus of its family. The sequence divergence between transposable elements and the consensus was corrected using the Jukes-Cantor distance, which corrects the divergence ($p$) by the formula $d = -3/4 \log(1-4/3p)$ (Jukes and Cantor 1969). For more details, see Supplemental Methods.

Identification of genomic rearrangements

The genome assemblies of V. dahliae strain JR2 and VdLS17 are available from NCBI under the assembly number GCA_000400815.2 and GCA_000952015.1, respectively (Faino et al. 2015). Whole-genome alignments between chromosomes of the genome assemblies of V. dahliae strains JR2 and VdLS17 were performed using MUMmer (Kurtz et al. 2004) and mined for genomic rearrangements and associated synteny breakpoints in V. dahliae strain JR2. Synteny breakpoints were further refined by mapping PacBio long-sequencing reads derived by genomic sequencing of V. dahliae strain JR2 to the genome of V. dahliae strain JR2 using BLASR (Chaisson and Tesler 2012), followed by manual refinement. To assess the presence or absence of genomic rearrangements in other V. dahliae strains, paired-end reads derived from genome sequencing (PRJNA169154) (de Jonge et al. 2012) were mapped onto the genome of V. dahliae strain JR2 using BWA (Li and Durbin 2010), and genomic regions surrounding the identified genomic rearrangements (±4 kb) were visually evaluated for the quantity of concordantly and discordantly mapped reads as well as orphan reads. For more details, see Supplemental Methods.

Identification and analyses of highly dynamic genomic regions

Whole-genome alignments between chromosomes of the complete genome assemblies of V. dahliae strains JR2 and VdLS17 (Faino et al. 2015) were performed using MUMmer (Kurtz et al. 2004). LS regions were manually defined by identifying regions accumulating alignment breaks and TEs. The presence/absence analysis of the AveI locus was performed by aligning paired-end reads from DNA sequencing of 11 V. dahliae strains (PRJNA169154, including JR2) (de Jonge et al. 2012) to the assembled genome of V. dahliae strain JR2 using BWA (Li and Durbin 2010). Raw read depth per genomic position was averaged per genomic window—window-size 5 kb, slide 500 bp (Fig. 4C), and window-size 500 bp, slide 100 bp (Supplemental Fig. S11), respectively —and subsequently G+C corrected as previously described (Yoon et al. 2009). Genomic reads of each individual additional V. dahliae strain were assembled using A5 pipeline (Tritt et al. 2012), and genome assemblies were subsequently aligned to the genome assembly of V. dahliae strain JR2 genome using MUMmer (Kurtz et al. 2004). For more details, see Supplemental Methods.

Single nucleotide polymorphisms (SNPs) were identified using GATK v2.8.1 (DePristo et al. 2011). SNPs derived from different strains were summarized in nonoverlapping windows of 1 kb, and the number of SNPs derived from individual strains was averaged per window. Absence of a SNP in a particular strain was only considered if the corresponding position displayed read coverage. For more details, see Supplemental Methods.

Competing interest statement

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

Acknowledgments

Work in the laboratory of B.P.H.J.T. is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands
Organization of Scientific Research (NWO). M.F.S. acknowledges the receipt of a VENI grant of ALW-NWO.

Author contributions: L.F. conceived the study, participated in its design and coordination, performed analyses, and contributed to writing of the manuscript; M.F.S. conceived the study, participated in its design and coordination, performed analyses, and wrote the manuscript; G.C.M.B. produced biological materials used in the study; X.S.K. and M.P. performed analyses; A.H.J.W. participated in the establishment of high-quality genome assemblies; B.P.H.J.T. conceived the study, participated in its design and coordination, and wrote the manuscript. All authors read and approved the final manuscript.

References

Amoyte SG, Tan XP, Pennerman K, Jimenez-Gasco MD, Klosterman SJ, Ma LJ, Dobyns KF, Versele P. 2012. Transposable elements in phytopathogenic Verticillium spp.: insights into genome evolution and inter- and in-species diversification. BMC Genomics 13: 314.

Bailey JA, Liu G, Eichler EE. 2003. An Ali transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 73: 823–834.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application to human genome assembly. BMC Bioinformatics 13: 138.

Chujo T, Scott B. 2014. Histone H3K9 and H3K27 methylation regulates chromatin formation and its regulation. PLoS Pathog 10: e1003947.

Cook DE, Mesarich CH, Thomma BP. 2015. Understanding plant immunity as a surveillance system to detect invasion. Annu Rev Phytopathol 53: 541–563.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet 10: 691–703.

Daverdin G, Rouxel T, Gout L, Aubert JN, Fudal I, Meyer M, Parlane F, Carpejaz F, Balesdent MH. 2012. Genome structure and reproductive behavior influence the evolutionary potential of a fungal phytopathogen. PLoS Pathog 8: e1003200.

de Jonge R, van Smeulendael K, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang Z, Usami T, Lievens B, Subbarao KV, et al. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc Natl Acad Sci 109: 11105–11110.

de Jonge R, Bolton MD, Kombrink A, van den Berg GC, Yadeta KA, Thomma BP. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res 23: 1271–1282.

De Pinto MA, Banks E, Poplin R, Gartempeh KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, et al. 2011. A framework for population discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–498.

Dong S, Raffenae S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Dev 35: 57–65.

Faino L, Seidl MF, Datema E, van den Berg GC, Janssen A, Wittenberg AH, de Jonge R, Bolton MD, Kombrink A, van den Berg GC, Yadeta KA, Thomma BP. 2013. Single-Molecule Real-Time sequencing combined with optical mapping yields completely finished fungal genomes. Mol Microbiol 92: 413–434.

Fradin EF, Thomma BP. 2006. Physiology and molecular aspects of Verticillium wilt diseases caused by V. dahiae and V. albo-atrum. Mol Plant Pathol 7: 71–86.

Fradin EF, Zhang Z, Ayala JC, Castorverde CD, Nazar RN, Robb J, Liu CM, Thomma BP. 2009. Genetic dissection of Verticillium wilt resistance mediated by tomato Ve1. Plant Physiol 150: 320–332.

Fudal I, Ross S, Brun H, Besnard AL, Ermel M, Kuhn ML, Balesdent MH, Rosenthal T. 2009. Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward virulence in Leptosphaeria maculans. Mol Plant Microbe Interact 22: 932–941.

Galakza JM, Freitag M. 2014. Variability of chromosome structure in pathogenic fungi— of ends and odds. Curr Opin Microbiol 20: 19–26.

Gijzen M. 2009. Runaway repeats force expansion of the Phytophthora infes- tans genome. Genome Biol 10: 241.

Goodwin SB, M’Barek SB, Dollison B, Wittenberg AH, Crane CF, Hane JK, Fossan AJ, Van der Lee DA, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat pathogen Mycosphaerella gramincola reveals dispensable structure, chromosome plasticity, and stealth pathogene- sis. PLoS Genet 7: e1002070.

Grandaubert J, Lowe RG, Soyer JL, Schoch CL, Van de Wouw AP, Fudal I, Robbertse B, Lapalu N, Links MG, Ollivier B, et al. 2014. Transposable element-assisted evolution and adaptation to host plant within the Leptosphaeria maculans-Leptosphaeria biglobosa species complex of fungal pathogens. BMC Genomics 15: 891.

Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Rafaela S, Torres-Alalibo T, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461: 393–398.

Hedges DJ, Deininger PL. 2007. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat Res 616: 46–59.

Heitman J, Kronstad JW, Taylor JW, Casselton L. 2007. Sex in fungi: molecular determination and evolutionary implications. ASM Press, Washington, DC.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12: 491.

Inderbitzin P, Subbarao KV. 2014. Verticillium systematics and evolution: how confusion impedes Verticillium wilt management and how to re- solve it. Phytopathology 104: 564–574.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In Mammalian protein metabolism, Vol. 3, pp. 131–132. Elsevier, Academic Press, New York.

Klosterman SJ, Atallah ZK, Vallad GE, Subbarao KV. 2009. Diversity, pathogenicity and management of Verticillium species. Annu Rev Phytopathol 47: 39–62.

Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BP, Chen Z, Henrisrat B, Lee YH, Park J, et al. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog 7: e1002137.

Krejci L, Altmanova V, Spirek M, Zhao X. 2012. Homologous recombina- tion and its regulation. Nucleic Acids Res 40: 5795–5818.

Kurtz S, Philippakis AA, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5: R12.

Lewis ZA, Honda S, Khalfallah TK, Jeffress JK, Freitag M, Mohr F, Schübler D, Weller MA. 2009. Retro EA: retro element analysis. Nucleic Acids Res 37: 574–578.

Lisch D. 2013. How important are transposons for plant evolution? Annu Rev Genet 47: 9–62.

Lisch D, Groenen EE, Bozic J, Pi-ta A, Dufresne M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in eukaryotic fungi following host jumps in the Irish potato famine pathogen lineage. Proc Natl Acad Sci 107: 10946–10951.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantification of short DNA sequences from RNA-seq data. Nat Biotechnol 26: 539–545.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantification of short DNA sequences from RNA-seq data. Nat Biotechnol 26: 539–545.

Nature 461: 393–398.

Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. 2000. A telomeric complex, DCDC. Mcm21, is required for proper telomere capping in vertebrates. Curr Opin Genet Dev 11: 166–172.

Pi-ta A, Dufresne M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in eukaryotic fungi following host jumps in the Irish potato famine pathogen lineage. Proc Natl Acad Sci 107: 10946–10951.

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RH, Zody MC, Kunjeti SG, Donofrio NM, et al. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science 330: 1540–1543.
Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bouzar S, et al. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat Commun* 2: 202.

Rovenich H, Boshoven JC, Thomma BP. 2014. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Curr Opin Plant Biol* 20C: 96–103.

Schrader L, Kim JW, Ence D, Zimin A, Klein A, Wyschetzki K, Weichselgartner T, Kemena C, Stökl J, Schulzner E. 2014. Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat Commun* 5: 5495.

Seidl MF, Thomma BP. 2014. Sex or no sex: Evolutionary adaptation occurs regardless. *Bioessays* 36: 335–345.

Seidl MF, Faino L, Shi-Kunne X, van den Berg GC, Bolton MD, Thomma BP. 2015. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. *Mol Plant Microbe Interact* 28: 362–373.

Selker EU. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu Rev Genet* 24: 579–613.

Song Y, Zhang Z, Seidl MF, Majer A, Jakse J, Javornik B, Thomma BP. 2016. Broad taxonomic characterization of *Verticillium* wilt resistance genes reveals ancient origin of the tomato Ve1 immune receptor. *Mol Plant Pathol* doi: 10.1111/mpp.12390.

Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent MH, Connolly LR, Freitag M, Rouxel T, et al. 2014. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet* 10: e1004227.

Soyer JL, Rouxel T, Fudal I. 2015. Chromatin-based control of effector gene expression in plant-associated fungi. *Curr Opin Plant Biol* 26: 51–56.

Thomma BP, Nünnert T, Joosten MH. 2011. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell* 23: 4–15.

Thomma BP, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JA, Faino L. 2015. Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genet Biol* 90: 24–30.

Thon MR, Pan H, Diener S, Papalas J, Taro A, Mitchell TK, Dean RA. 2006. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol* 7: R16.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111.

Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. An integrated pipeline for de novo assembly of microbial genomes. *PLoS One* 7: e42304.

van Hooft JJ, Snel B, Seidl MF. 2014. Small homologous blocks in *Phytophthora* genomes do not point to an ancient whole-genome duplication. *Genome Biol Evol* 6: 1079–1085.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating γ-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* 8: 77–80.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8: 973–982.

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

Wicker T, Oberhaensli S, Parlarge F, Buchmann JP, Shatalina M, Roffler S, Ben-David R, Doležel J, Šimková H, Schulze-Lefert P, et al. 2013. The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nat Genet* 45: 1092–1096.

Yoon ST, Xuan ZY, Makarov V, Ye K, Sebat J. 2009. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* 19: 1586–1592.

Zhang Z, Song Y, Liu CM, Thomma BP. 2014. Mutational analysis of the Ve1 immune receptor that mediates *Verticillium* resistance in tomato. *PLoS One* 9: e99511.

Received January 30, 2016; accepted in revised form June 17, 2016.