Supporting Online Material for:

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

Luigi Faino, Michael F Seidl, Xiaoqian Shi-Kunne, Marc Pauper, Grardy CM van den Berg, Alexander HJ Wittenberg, and Bart PHJ Thomma*****

***To whom correspondence should be addressed. E-mail: bart.thomma@wur.nl**

This PDF file includes: Supplemental Methods, and Supplemental Figures 1-14

Supplemental Methods

Genome annotation and dynamics

Gene predictions for the whole-genome assembly of *V. dahliae* strain JR2 (Faino et al. 2015) was performed using the Maker2 software (Holt and Yandell 2011). To this end, RNA-seq reads derived from different *in vitro* and *in planta* conditions (de Jonge et al. 2012) were mapped to the genome using Tophat2 (default settings) (Trapnell et al. 2009). Additionally, gene sequences derived from previous genome annotations as well as protein sequences from 35 fungal proteomes were used as additional evidence (Klosterman et al. 2011; de Jonge et al. 2013; Seidl et al. 2015).

Homology between 13 fungal species was assessed using OrthoMCL (default settings) (Li et al. 2003). Genome sequences and annotation for *Fusarium oxysporum* f. sp. *lycopersici*, *Colletotrichum higginsianum* were retrieved from the Ensembl fungi database (http://fungi.ensembl.org/). The data for *Acremonium alcalophilum* were downloaded from JGI database, and the data for *V. alfalfa* MS102 (Klosterman et al. 2011) were obtained from the *Verticillium* database at the Broad Institute. Sequence similarity between proteins was established by all-vs.-all analyses using BLASTp (E-value cutoff 1e-5, soft filtering)(Altschul et al. 1990). Orthologous gene pairs between fungal species, and paralogous gene pairs within *V. dahliae* strain JR2, were extracted from the OrthoMCL gene families. Ks values between gene pairs, as defined by OrthoMCL families, were calculated using the Nei-Gojobori algorithm included in the K_aK_s_Calculator 2.0 package (Wang et al. 2010). The coding sequences of gene pairs were aligned using protein alignment as a guide. The phylogenetic tree was generated by RAxML (Stamatakis 2006) using concatenated protein sequences of 3,492 single-copy orthologs that are conserved among eight fungal species, where only a single representative species was chosen for the fungal genera *Colletotrichum* and *Fusarium* (Supplemental Fig. 14).

Repetitive elements were identified as described in Faino et al. (2015). Briefly, repetitive elements were identified using RepeatScout, LTR_Finder and LTRharvest, and the repetitive elements identified by the different software were combined (non-redundant). Repetitive elements were further classified as described by Wicker et al. (2007). Open reading frames within the transposable elements were identified by BlastN and BlastX (Camacho et al. 2009) searches against NCBI NR databases as well as by InterProScan (Jones et al. 2014). Repetitive elements that could not be

classified were defined as ‘unknown’. 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). Reads were mapped onto the genome assembly of *V. dahliae* strain JR2 using Tophat2 (default parameters, max-intron length 1,000 nt) (Trapnell et al. 2009). Mapped reads were summarized using the R package GenomicAlignments (‘summarizeOverlaps’) (Lawrence et al. 2013), and the expression per repetitive element, excluding simple repeats and repeats overlapping genes, was reported as Reads Per Kilobase of transcript per Million mapped reads (RPKM). To address reads that map multiple-times in the genome, and their influence on the expression estimates for repetitive elements, we filtered reads mapped with Tophat2 based on their mapping quality. Only reads with mapping quality >5 were retained, summarized and reported as RPKM. To estimate gene expression, mapped reads (Tophat2; default parameters, max-intron length 1,000 nt) were summarized, and reported as RPKM.

To estimate divergence time of transposable elements, each individual copy of a transposable element was aligned to the consensus of its family using needle, which is part of the EMBOSS package (Rice et al. 2000). The consensus sequence for each transposable element family was determined by performing multiple-sequence alignment of all copies belonging to the same family using mafft (Katoh and Standley 2013) (each individual copy needed to be longer than 400 bp). For the consensus sequence, only columns with > 1 aligned sequence (excluding gaps) were considered, for which the nucleotide occurring in the majority of sequences was used for the consensus sequence (ties, a nucleotide randomly chosen from the tie was picked). The sequence divergence between transposable elements and the consensus was corrected using the Jukes-Cantor distance, which corrects the divergence (p) by the formula $d=-3/4\log_e(1-4/3p)$(Jukes and Cantor 1969).

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 (Faino et al. 2015) were performed using nucmer (default settings), which is part of the MUMmer 3.0 package(Kurtz et al. 2004). To remove spurious hits, the alignments were

subsequently filtered by length, retaining alignments > 15 kb and 99% identity. These parameters were chosen based on the average nucleotide identity between the two *V. dahliae* strains (99.98%), as well as the average length of unique sequences in the genome. These whole-genome alignments were further mined for genomic rearrangements and associated synteny breakpoints in *V. dahliae* strain JR2. The identified synteny breakpoints were further refined by mapping PacBio long-sequencing reads derived by genomic sequencing of *V. dahliae* strain VdLS17 to the genome of *V. dahliae* strain JR2 using Blasr (default settings) (Chaisson and Tesler 2012), followed by manual refinement. The alignments were plotted using Sushi R package (Phanstiel et al. 2014). GEvo (Lyons et al. 2008) was used to identify syntenic regions between *V. dahliae* strains JR2 and VdLs17, where only gene-coding regions were used as anchors between the syntenic chromosomal regions.

To assess the presence or absence of genomic rearrangements in other *V. dahliae* strains, paired-end reads derived from genome sequencing (de Jonge et al. 2012) (PRJNA169154) were mapped onto the genome of *V. dahliae* strain JR2 using BWA (BWA-mem algorithm) (Li and Durbin 2010). 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.

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 nucmer (settings: -maxmatch), which is part of the MUMmer 3.0 package (Kurtz et al. 2004). Alignments separated by gaps <500 bp were merged in unique and contiguous alignments. LS regions were manually defined by identifying regions accumulating alignments breaks and TEs. Additionally, genomic reads derived from ten *V. dahliae* strains (de Jonge et al. 2012) were mapped on the *V. dahliae* strain JR2 and VdLs17, respectively, using Bowtie2 (default settings) (Langmead et al. 2009). The genomic coverage was determined by BEDtools (Quinlan and Hall 2010). The phylogenetic tree of eleven different *V. dahliae* strains was generated using RealPhy (Bertels et al. 2014) using either *V. dahliae* strain JR2 or VdLs17 as a reference strain.

The presence/absence analysis of the *Ave1* locus was performed by aligning paired-end reads from DNA sequencing of eleven *V. dahliae* strains (de Jonge et al.

2012) (including JR2) to the assembled genome of *V. dahliae* strain JR2 using BWA (BWA-mem algorithm) (Li and Durbin 2010). Raw read depth per genomic position was averaged per genomic windows (window-size 5 kb; slide 500 bp (Figure 3D; E) and window-size 500 bp; slide 100 bp (Supplemental Fig 10B), respectively), and subsequently performed a G+C correction similarly as previously described (Yoon et al. 2009). Briefly, we adjusted these averaged raw read depth (ARD) based on the observed deviation of read depth for a given G+C percentage. To this end, we first determined the average ARD for G+C percentages ranging from 0-100% (by 1%). Subsequently, we corrected the ARD using the formula $ARDc_i = ARD_i * (m/m_{G+C})$, where $ARDc_i$ is the corrected ARD in the *i*th window, ARD_i is the ARD in the *i*th window, m is the average ARD over all windows, and m_{G+C} is the average ARD for all windows with the same G+C percentage as the *i*th window (Yoon et al. 2009). Additionally, the genomic reads of each individual additional *V. dahliae* strain were assembled using A5 pipeline (v.20140113) (Tritt et al. 2012). The assembled genomes were aligned to the genome assembly of *V. dahliae* strain JR2 genome using nucmer (settings: -maxmatch), which is part of the MUMmer 3.0 package (Kurtz et al. 2004). Overlaps between genomic coordinates of different genome features, e.g. lineage-specific regions, genes or transposable elements, were assessed by BEDtools (v2.24.0) (Quinlan and Hall 2010) or by the R package GenomicRanges (Lawrence et al. 2013).

Single nucleotide polymorphisms (SNPs) were identified using GATK v2.8.1 (DePristo et al. 2011). Briefly, paired-end reads derived from ten *V. dahliae* strains (de Jonge et al. 2012) were mapped onto the complete genome assembly of *V. dahliae* strain JR2 (Faino et al. 2015) using BWA (BWA-mem algorithm) (Li and Durbin 2010). Using GATK v2.8.1 (DePristo et al. 2011), mapped reads were locally realigned to minimize the number of mismatches over all reads, and subsequently genomic variants (SNPs) were called using GATK's UnifiedGenotyper (default settings; emitting threshold 20, haploid organism) and resulting variants were quality filtered (quality > 50; phred-scaled quality score for the assertion), depth > 10 and allelic frequency > 0.9). SNPs derived from different strains were summarized in non-overlapping windows of 1 kb, and the number of SNPs derived from the individual strains were averaged per window. Absence of a SNP in a particular strain was only considered if the corresponding position displayed read coverage.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol Biol Evol* **31**: 1077-1088.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST plus : architecture and applications. *BMC Bioinformatics* **10**: 421.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**: 238.

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

de Jonge R, van Esse HP, 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 USA* **109**: 5110-5115.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**: 491-498.

Faino L, Seidl MF, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, Thomma BPHJ. 2015. Single-Molecule Real-Time sequencing combined with optical mapping yields completely finished fungal genomes. *mBIO* **6**: e00936-00915.

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

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. *Mammalian protein metabolism* **3**: 21-132.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

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

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biology* **5**: R12.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comp Biol* **9**: e1003118.

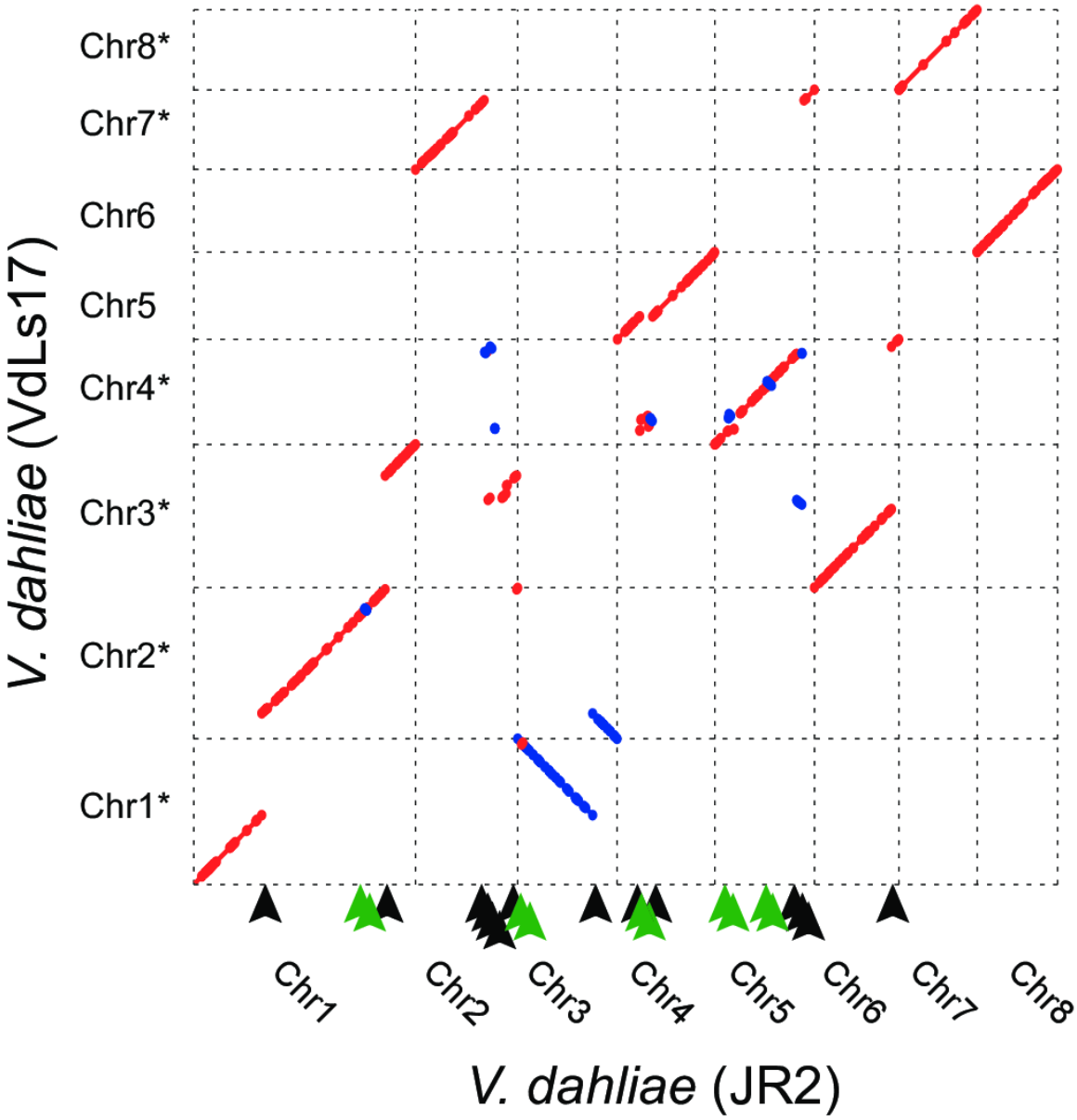
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589-595.

Li L, Stoeckert CJ, Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**: 2178-2189.

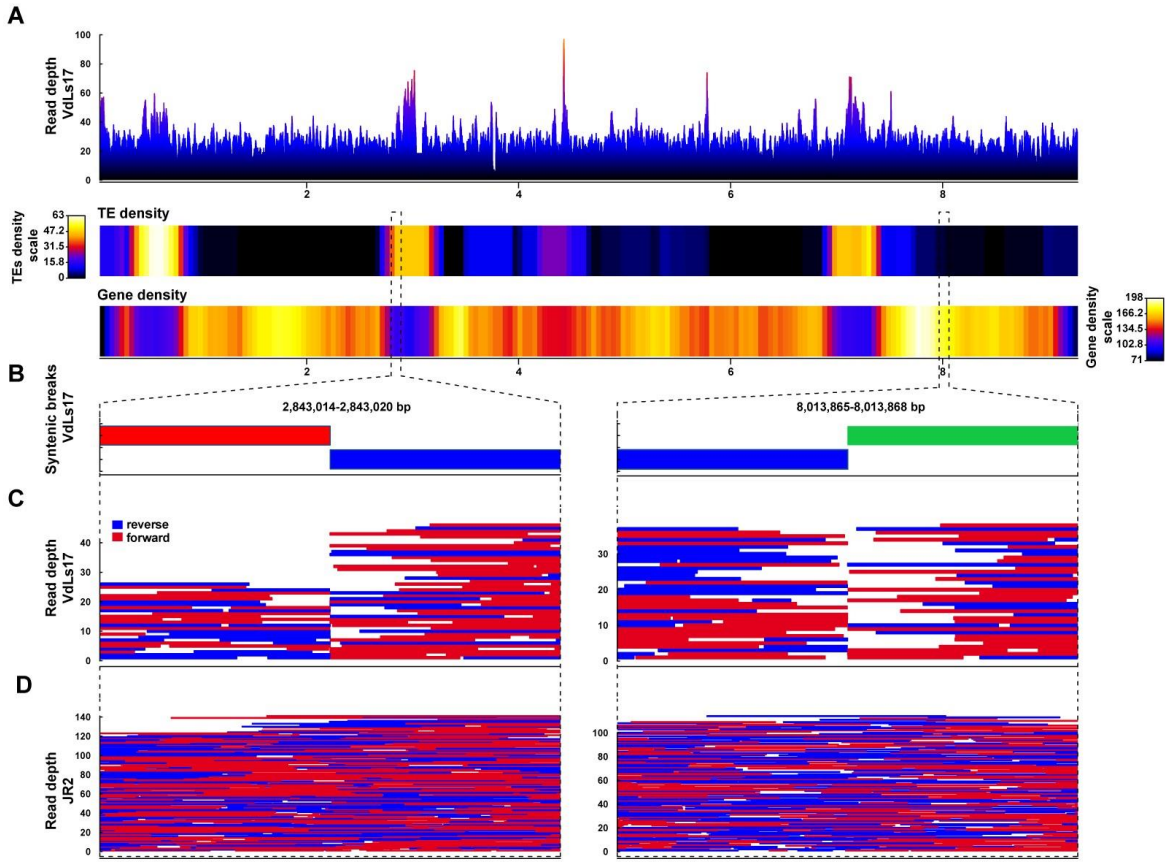
Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang HB, Wang XY, Bowers J, Paterson A, Lisch D et al. 2008. Finding and Comparing Syntenic Regions among *Arabidopsis* and the Outgroups Papaya, Poplar, and Grape: CoGe with Rosids. *Plant Physiol* **148**: 1772-1781.

Phanstiel DH, Boyle AP, Araya CL, Snyder MP. 2014. Sushi. R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**: 2808-2810.

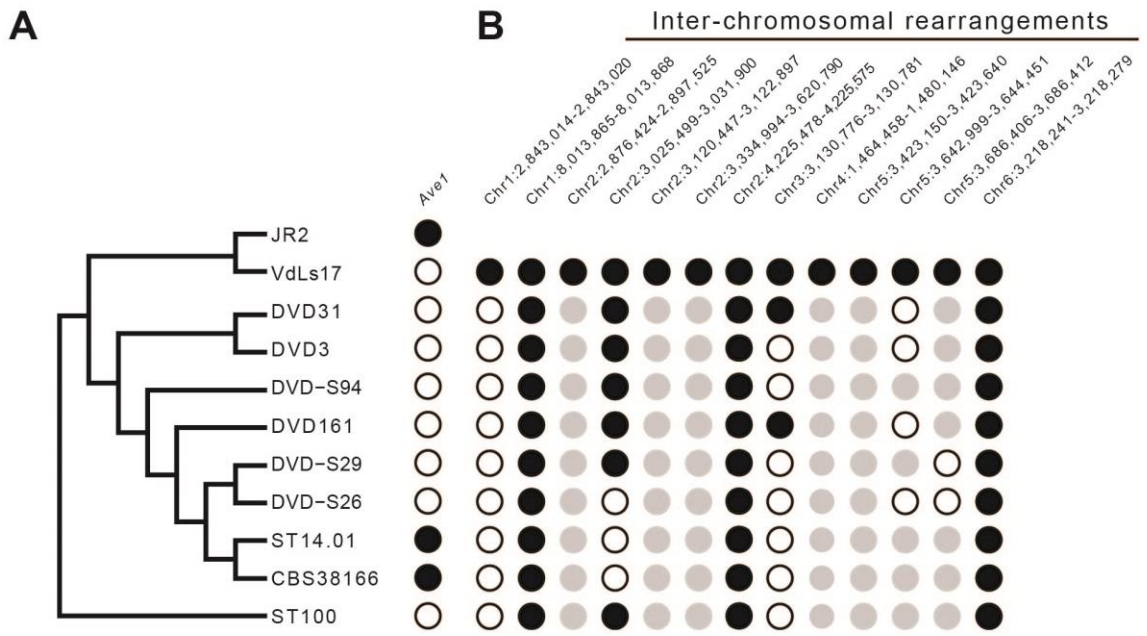
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European molecular biology open software suite. *Trends Genet* **16**: 276-277.
Seidl MF, Faino L, Shi-Kunne X, van den Berg GC, Bolton MD, Thomma BPHJ. 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.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.
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.
Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-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. *Nature Reviews Genetics* **8**: 973-982.



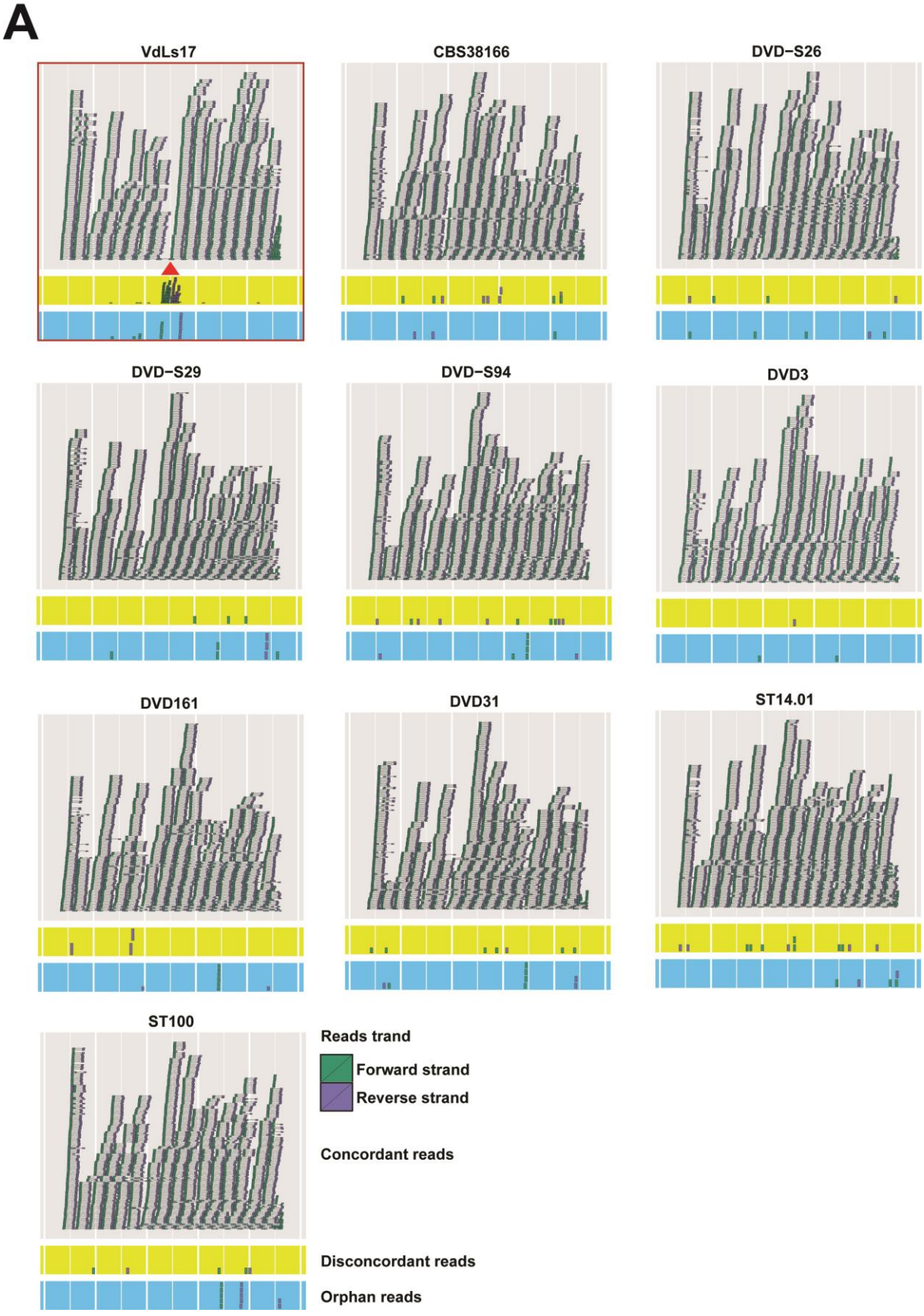
Supplemental Fig. S1 – Whole-genome alignment between two *Verticillium dahliae* strains reveals frequent chromosomal rearrangements. The whole-genome dot-plot displays structural polymorphisms between chromosomes of *V. dahliae* strains JR2 and VdLs17. Forward alignments are shown in red, reverse alignments in blue. Identified synteny breakpoints in *V. dahliae* strain JR2 are indicated by triangles (black for inter-chromosomal, green for intra-chromosomal). For clarity, particular chromosomes of *V. dahliae* strain VdLs17 have been inverted as indicated by an asterisk.

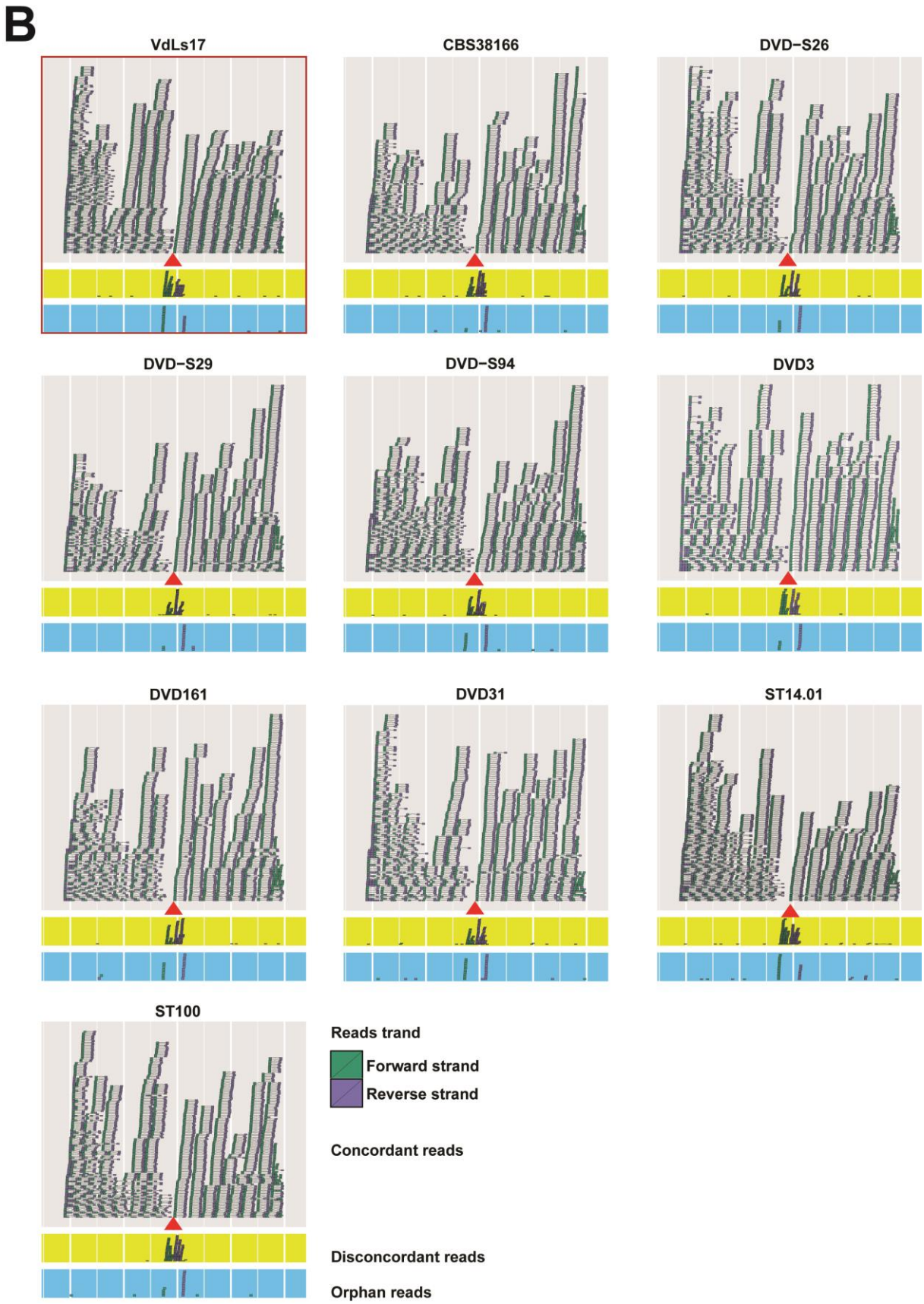


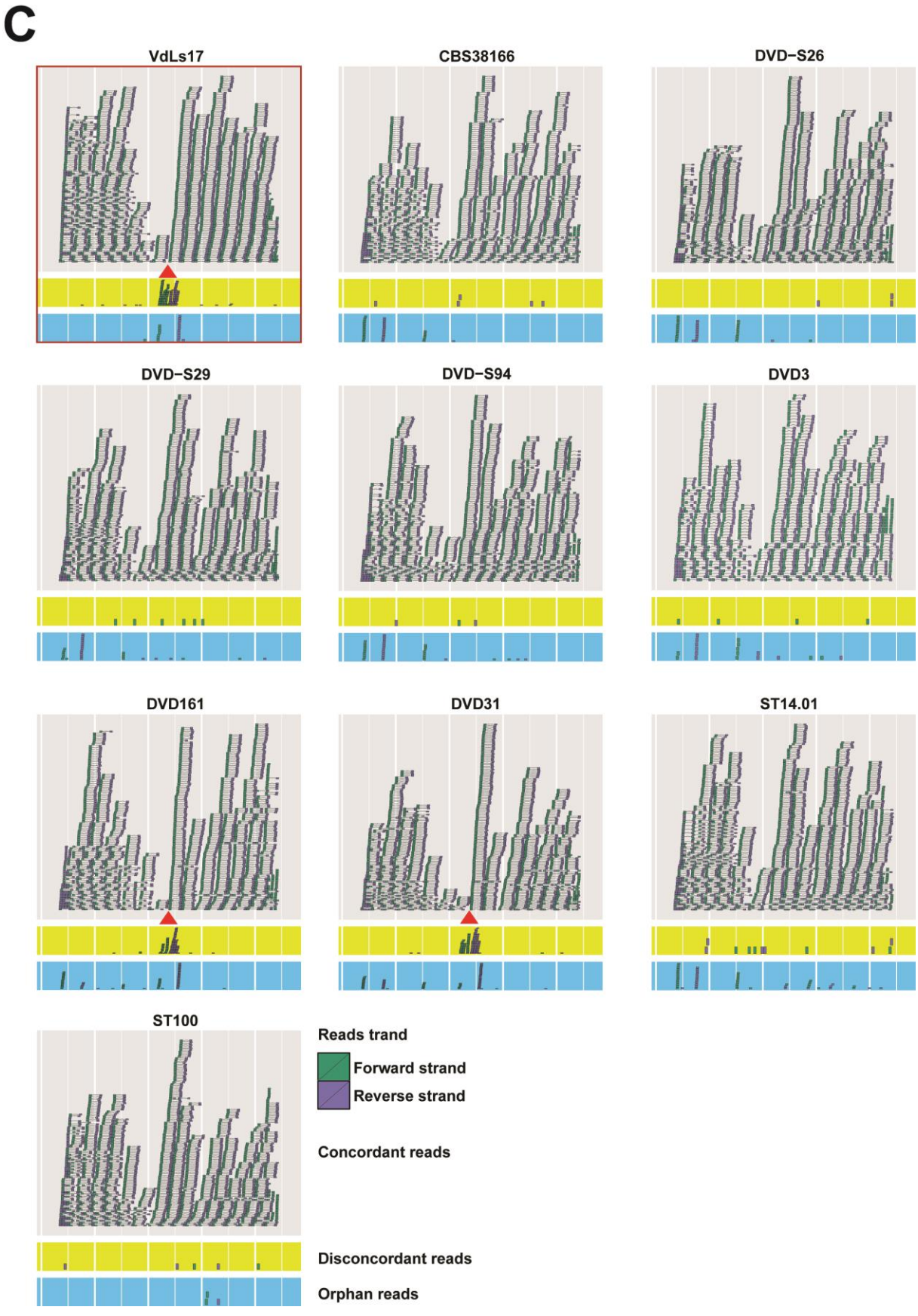
Supplemental Fig. S2 – Long-read mapping for high-resolution identification of synteny breakpoints. (A) High-resolution analysis of synteny breakpoints is exemplified for chromosome 1 of *V. dahliae* strain JR2. Long reads derived from genomic sequencing of *V. dahliae* strain VdLs17 are mapped onto chromosome 1 of *V. dahliae* strain JR2. The first 50 kb of chromosome 1 are omitted due to high read depth resulting from the collapse of the ribosomal gene cluster. The heatmaps indicate the transposable element (TE) density and the gene density, respectively, using a color scale from white (high density) to black (low density). (B) Synteny breakpoints between *V. dahliae* strain JR2 and VdLs17 chromosomes (colors of the contigs refer to Fig. 1). (C) Mapping of long reads derived from *V. dahliae* strain VdLs17 identifies synteny breakpoints in high resolution. Blue lines indicate long reads mapped on the forward strand, while red lines indicate reads mapped on the reverse strand. (D) Mapping of long reads derived from *V. dahliae* strain JR2 onto chromosome 1 of *V. dahliae* strain JR2 displays continuous mapping over the breakpoints.



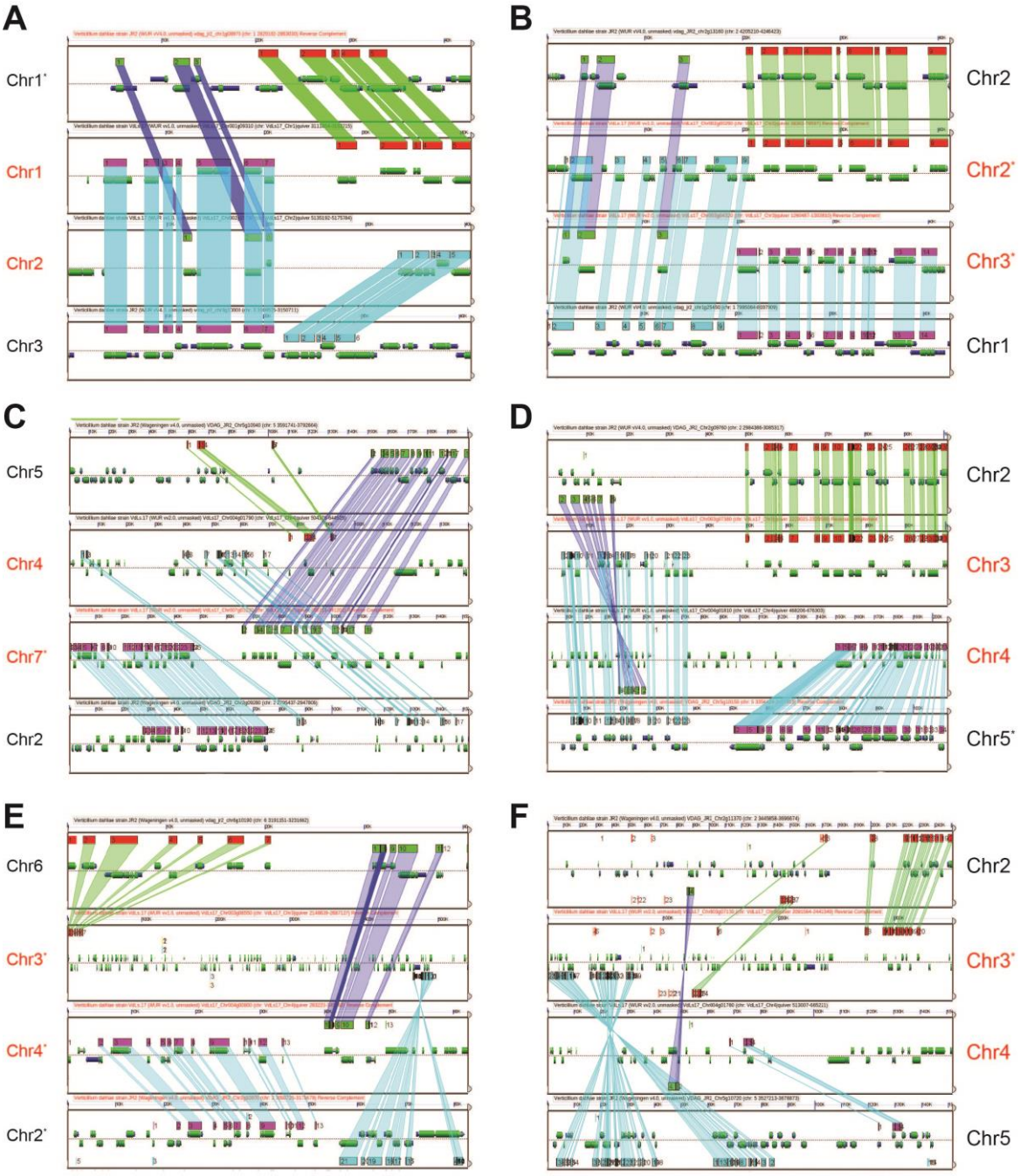
Supplemental Fig. S3 – The *Ave1* locus is conserved, yet not monophyletic, in the *Verticillium dahliae* population. (A) The unrooted phylogenetic tree was constructed with RealPhy using *V. dahliae* strain JR2 as a reference and paired-end reads from 10 additional *V. dahliae* strains. The presence (filled) or absence (empty) of the *Ave1* locus in the respective strain is indicated by circles. (B) The presence (black) or absence (white) of inter-chromosomal rearrangements, revealed by paired-end read analyses (Supplemental Fig 4), is indicated in the respective strains by circles. Grey circles indicate situations in which the exact constitution of the synteny breakpoint in the other *V. dahliae* strains could not be determined due to lineage-specific regions and/or repetitive elements.


A
VdLs17
CBS38166
DVD−S26
DVD−S29
DVD−S94
DVD3
DVD161
DVD31
ST14.01
ST100
Reads trand
Forward strand
Reverse strand
Concordant reads
Disconcordant reads
Orphan reads

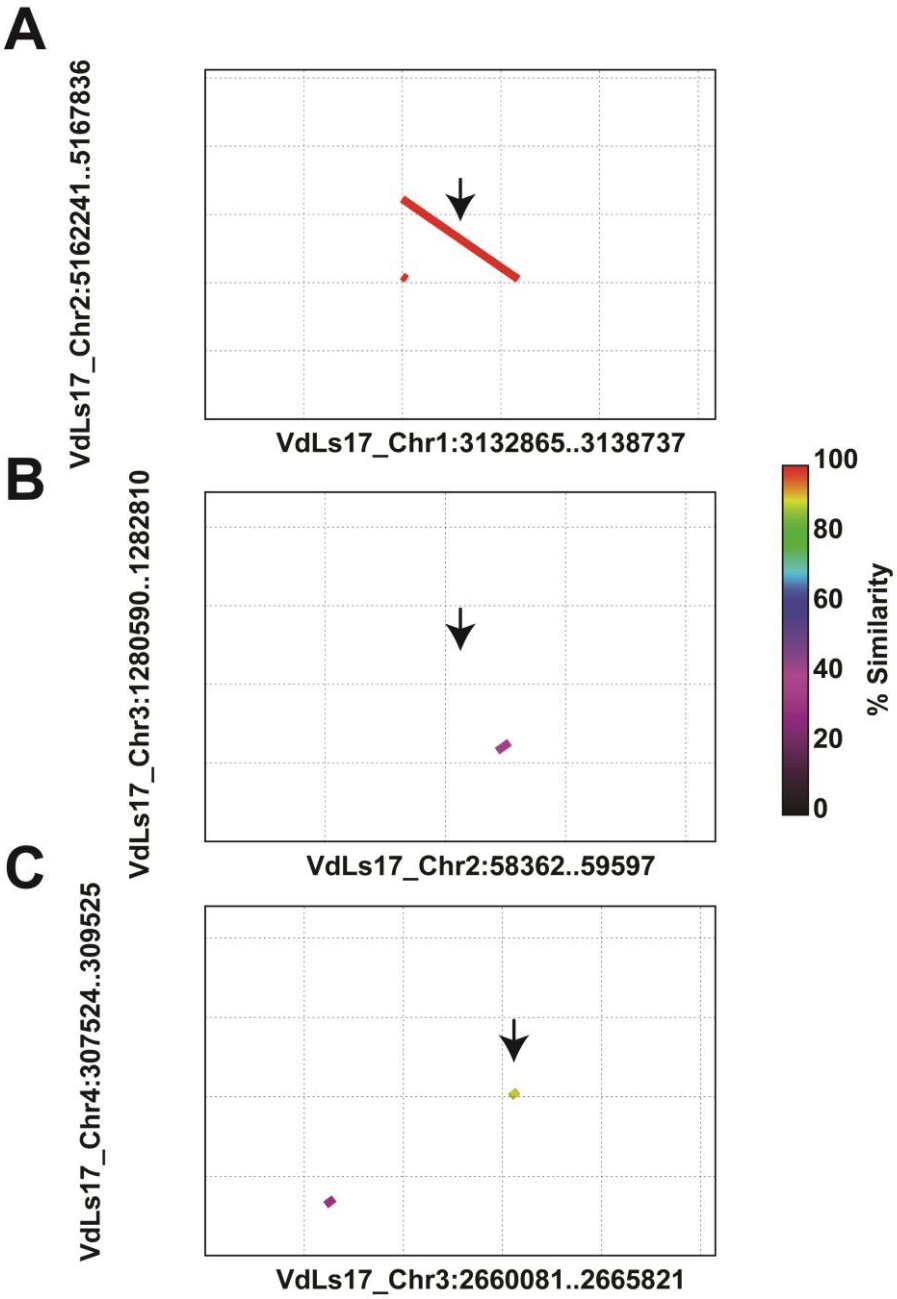

B
VdLs17
CBS38166
DVD−S26
DVD−S29
DVD−S94
DVD3
DVD161
DVD31
ST14.01
ST100
Reads trand
Forward strand
Reverse strand
Concordant reads
Disconcordant reads
Orphan reads


C
VdLs17
CBS38166
DVD−S26
DVD−S29
DVD−S94
DVD3
DVD161
DVD31
ST14.01
ST100
Reads trand
Forward strand
Reverse strand
Concordant reads
Disconcordant reads
Orphan reads

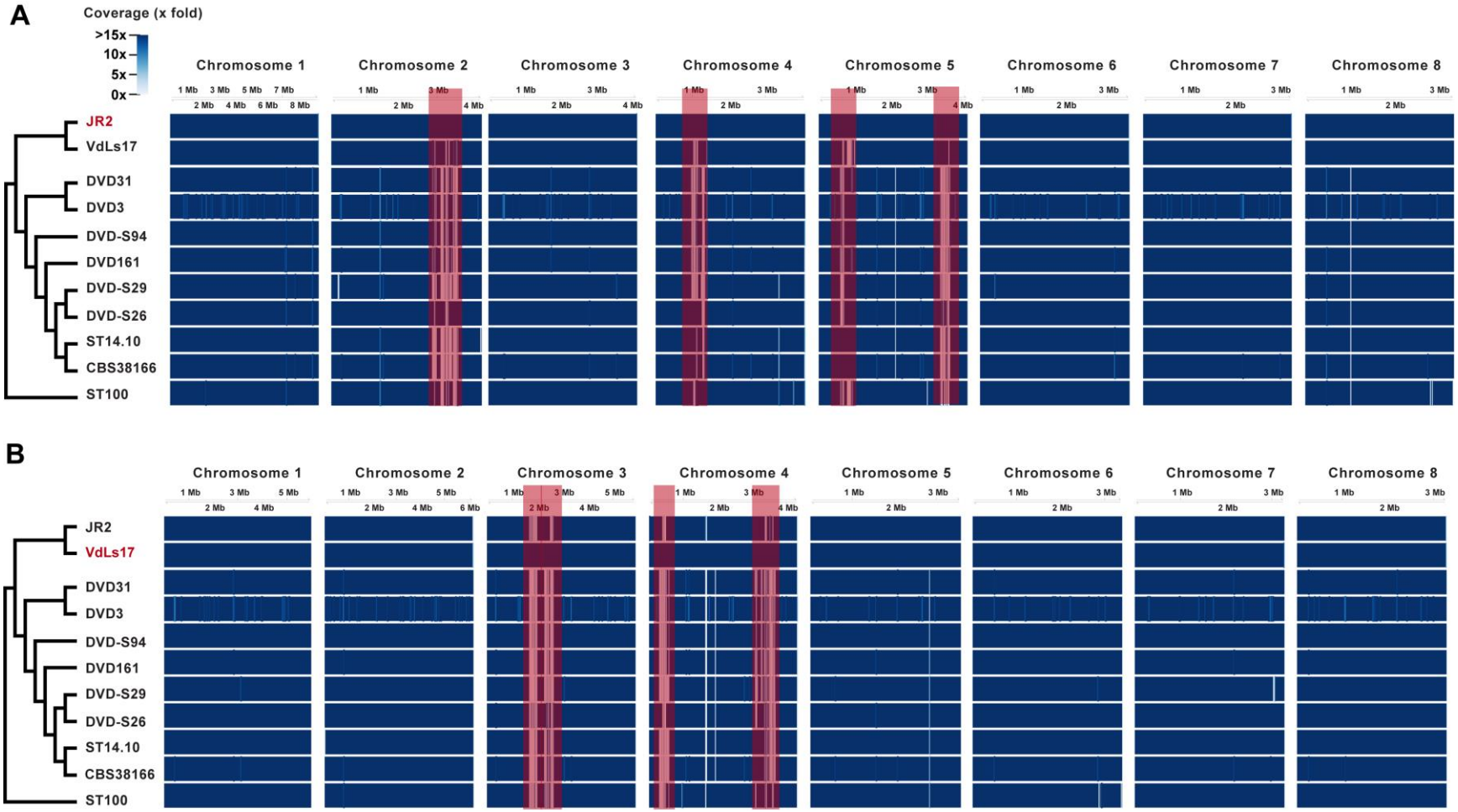
Supplemental Fig. S4 – Paired-end read analyses reveal distinct rearrangement patterns among *V. dahliae* strains. Paired-end reads derived from genomic sequencing of ten *V. dahliae* strains were mapped onto the genome assembly of *V. dahliae* strain JR2. Genomic regions surrounding three identified synteny breakpoints are highlighted, and concordant, discordant and orphan reads are displayed. A red box highlights the situation in *V. dahliae* strain VdLs17. The first synteny breakpoint (A) is only observed in *V. dahliae* strain VdLs17 and not in any other strain, the second synteny breakpoint (B) is present in all analyzed *V. dahliae* strains, and (C) the third is present in part of the *V. dahliae* population.



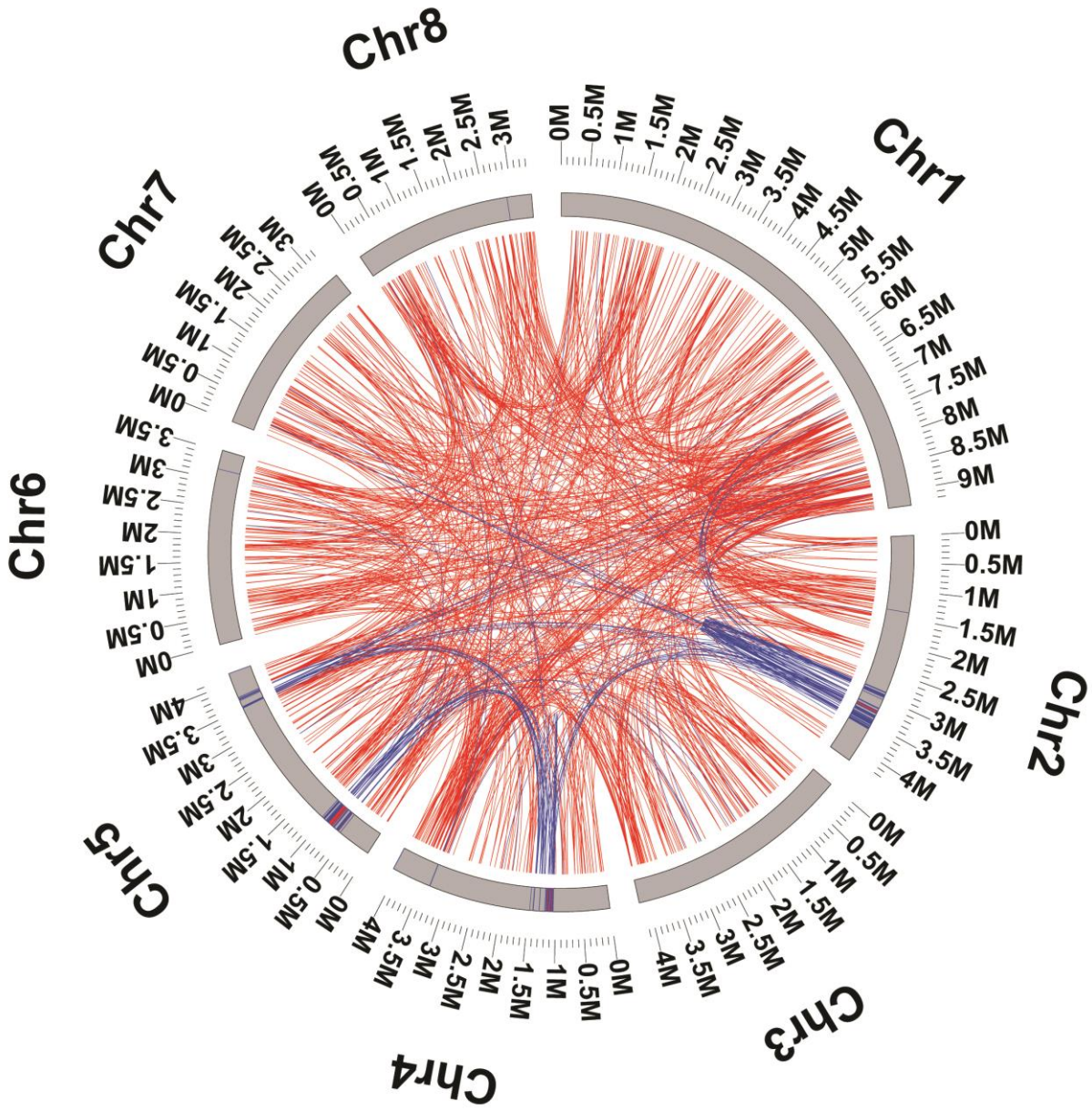
Supplemental Fig. S5 - Reconstruction of genomic translocations between *V. dahliae* chromosomes. Screenshot of the GEvo output for syntenic regions between chromosomes of *V. dahliae* strains JR2 (labeled in black) and VdLs17 (labeled in red). Each box represents a particular rearrangement between the two strains. Each panel (A to E, see Table 1 for details) contains 4 boxes that represent genomic regions of two chromosomes of both *V. dahliae* strains JR2 and VdLs17. Gene models are drawn as colored arrows for both DNA strands, with CDS colored green, RNA blue, and the full gene model as grey to show introns. GEvo was used to find regions of similarity (i.e. BLAST hits) between all pairwise sequence comparisons and are drawn as colored blocks above or below gene models, with each color representing one pairwise comparison. Ribbons between boxes connect syntenic regions. Only coding regions were used to identify syntenic blocks between different genomic regions.



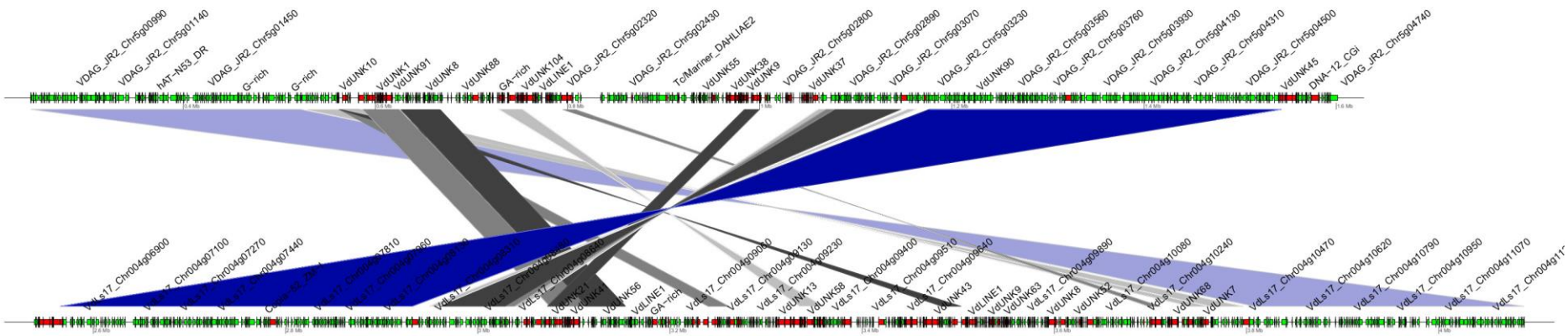
Supplemental Fig. S6 – Similarity between rearranged genomic location depicted in Fig 1B. Each plot represents a zoom into a rearranged region showed in Fig 1. The top panel indicates the synteny between regions showed in Fig 1B for the rearrangement 1, the middle panel indicates the synteny between genomic regions showed in Fig 1B for the rearrangement 2, while the bottom panel indicates the synteny between genomic regions showed in Fig 1B for the rearrangement 13. The syntenic regions are color coded based on the identity between the syntenic regions. the regions analyzed are the regions where the rearrangement occurred ± 5 kb from the rearranged point



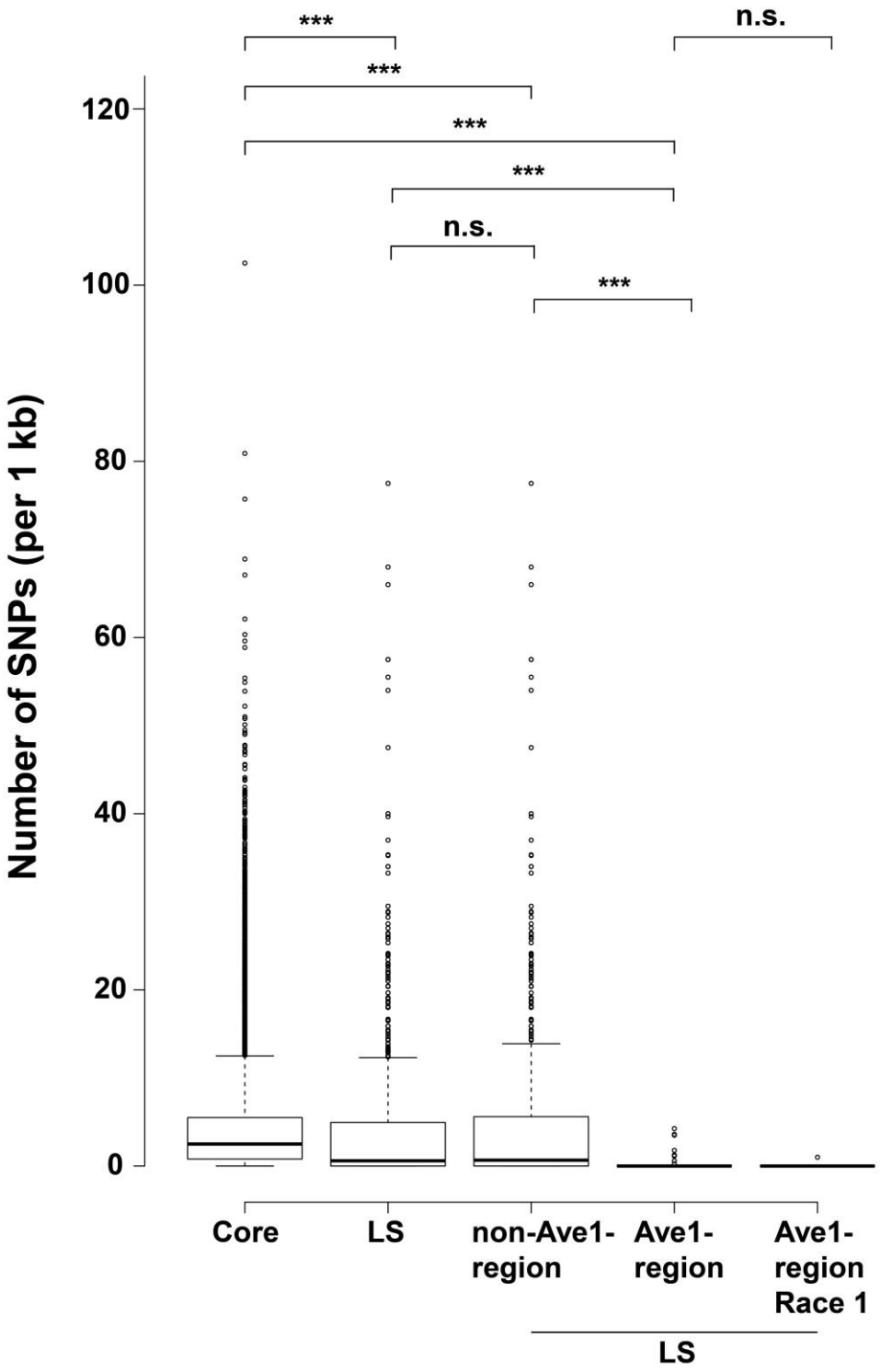
Supplemental Fig. S7 - Population structure and composition of lineage-specific regions in *Verticillium dahliae* strains JR2 (A) and VdLs17 (B). The unrooted phylogenetic tree was constructed with RealPhy using either *V. dahliae* strain JR2 (a) or *V. dahliae* strain VdLs17 (b) as a reference and paired-end reads from 10 additional *V. dahliae* strains. The coverage plot is derived from paired-end reads of 10 *V. dahliae* strains over the eight chromosomes of *V. dahliae* strain JR2 (A) or VdLs17 (B). Each row represents the genome coverage over the eight chromosomes. Blue blocks represent genome coverage higher than 15x while white regions show non-covered regions. LS regions identified in *V. dahliae* strain JR2 (A) or VdLs17 (B) are highlighted in red.



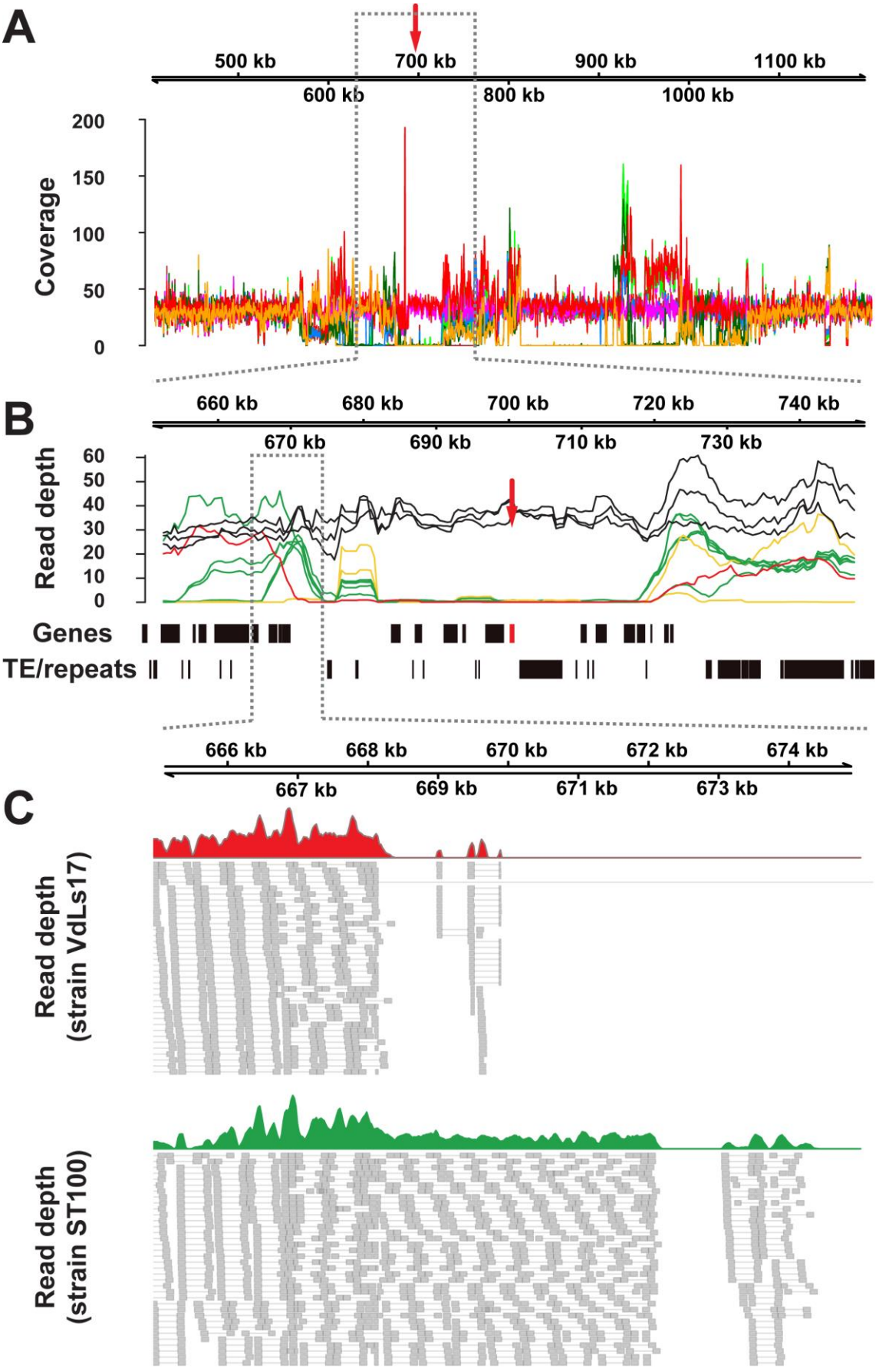
Supplemental Fig. S8 – Paralogous gene pairs in *V. dahliae* strain JR2. Circos diagram illustrating paralogous gene pairs in *V. dahliae* strain JR2. Paralogous gene pairs of which at least one member is located at the LS regions are connected with blue lines, while paralogous gene pairs located in the core genome are connected with red lines.



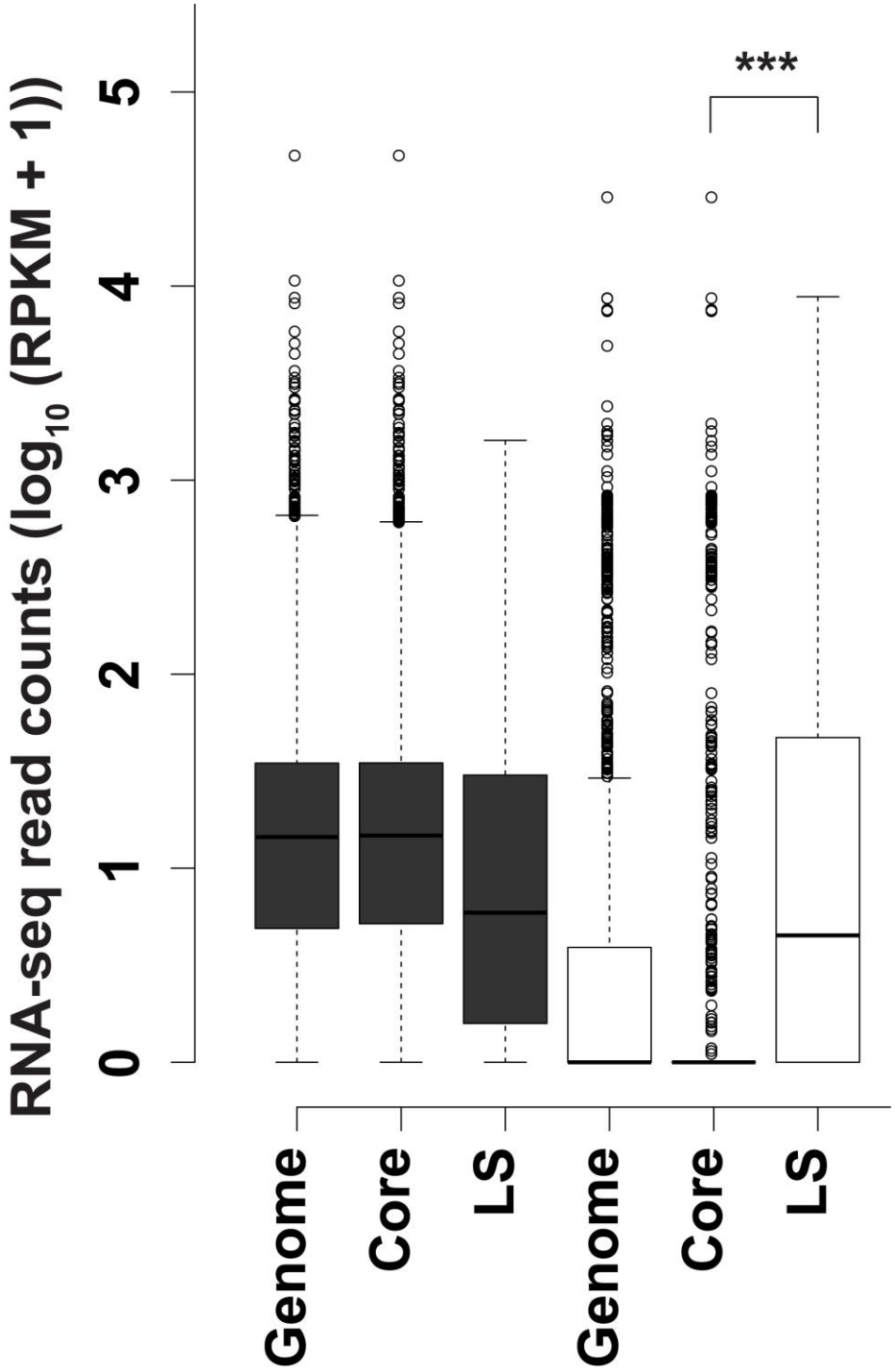
Supplemental Fig. S9 – Example of synteny and gene loss between LS regions in *V. dahliae* strain JR2 and VdLs17. Part of chromosome 5 of *V. dahliae* strain JR2 (top) is syntenic to part of chromosome 4 of *V. dahliae* strain VdLs17 (bottom). Syntenic non-LS regions between the two strains are indicated with blue ribbons, while syntenic regions within LS regions are indicated with grey ribbons. Genes are indicated in green and transposable elements are indicated in red.



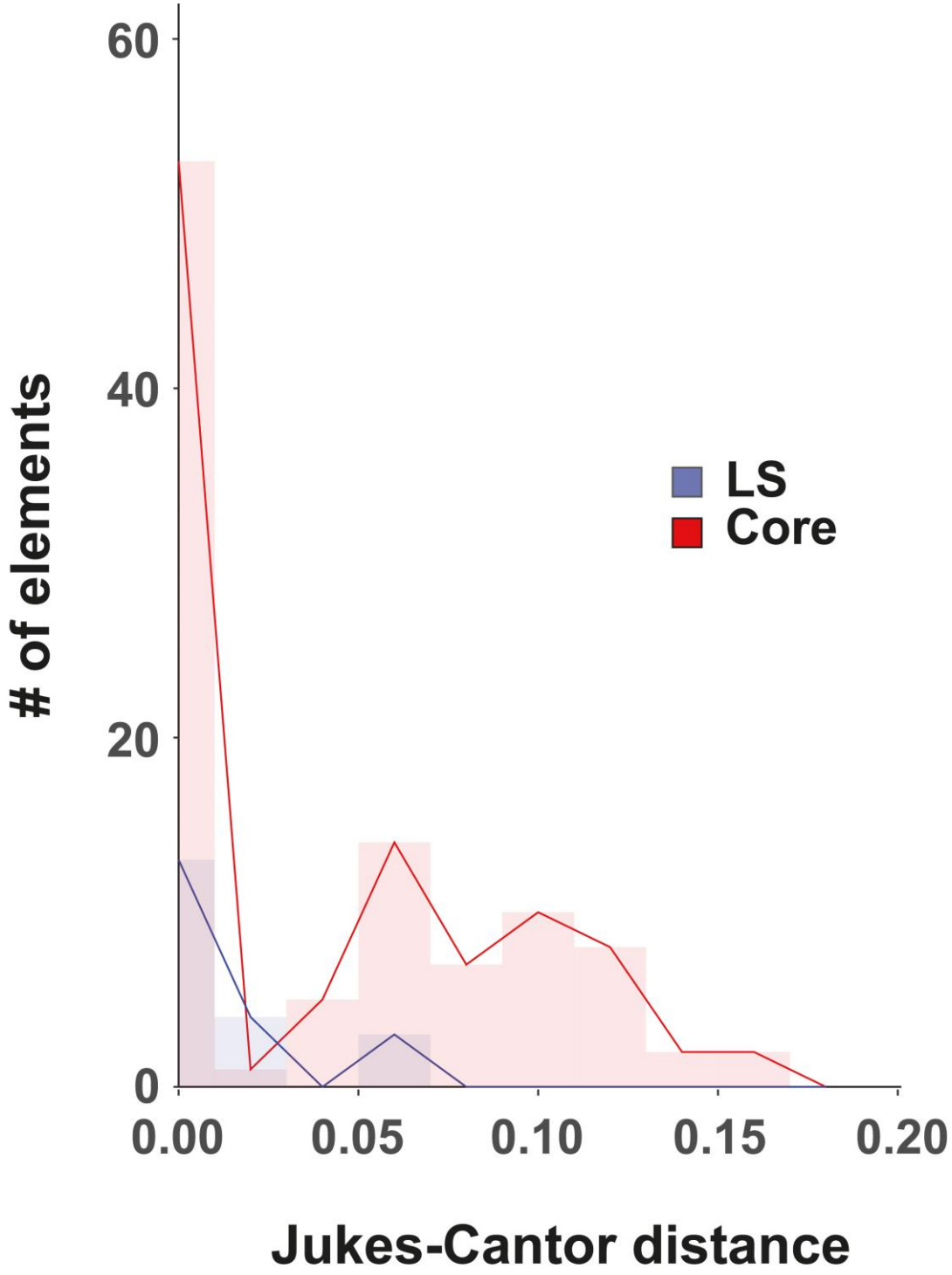
Supplemental Fig. S10 – SNP distribution across the *V. dahliae* JR2 genome. The box plot represents the average number SNPs derived in genomic windows of 1 kb between *V. dahliae* strain JR2 and 10 additional *V. dahliae* strains. The *Ave1* region was investigated using only race 1 strains. Statistical differences were assessed using a Wicoxon rank sum test.


A
500 kb
600 kb
700 kb
800 kb
900 kb
1000 kb
1100 kb
Coverage
200
150
100
50
0
B
660 kb
670 kb
680 kb
690 kb
700 kb
710 kb
720 kb
730 kb
740 kb
Read depth
60
50
40
30
20
10
0
Genes
TE/repeats
C
666 kb
667 kb
668 kb
669 kb
670 kb
671 kb
672 kb
673 kb
674 kb
Read depth (strain VdLs17)
Read depth (strain ST100)

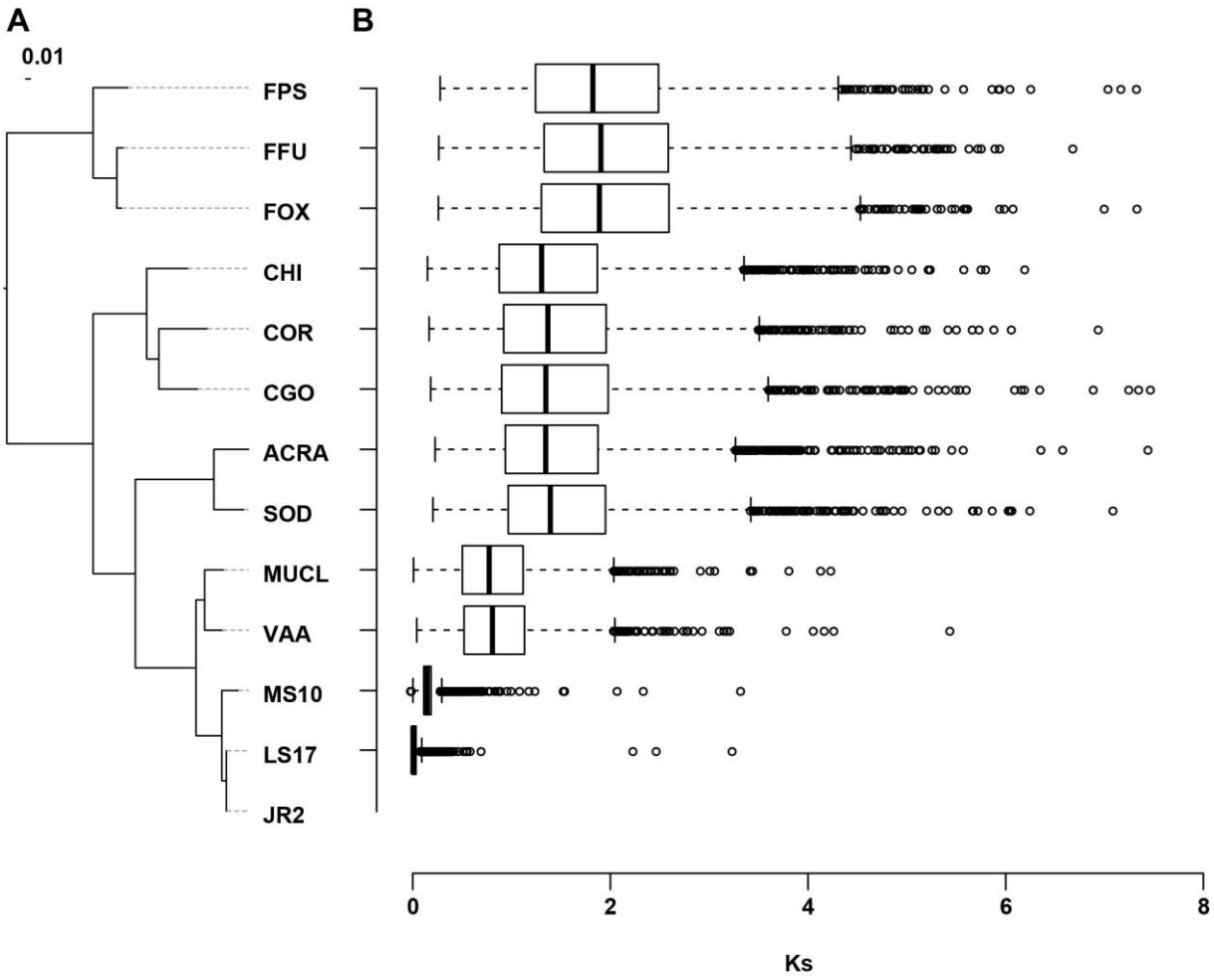
Supplemental Fig. S11 – *Verticillium dahliae* race 2 strains lack the *Ave1* locus. (A) Read depth was estimated by mapping the genomic reads derived from race 1 and race 2 strains to the *V. dahliae* strain JR2 reference (window 500 bp; slide 100 bp). The position of *Ave1* is indicated with a red arrow. (B) Magnification in of the *Ave1* locus, highlighting the four dominating coverage patterns. Lines indicate the corrected average read depth (per 5 kb window, 500 bp slide) of paired-end reads derived from genomic sequencing of eleven *V. dahliae* strains. Different colors indicate distinct patterns of coverage across the *Ave1* locus. Genes and transposable elements/repeats (excluding simple repeats) are indicated. (C) Detailed view of paired-end reads derived from *V. dahliae* strains VdLs17 and St100, revealing distinct regions where the sequence coverage drops.



Supplemental Fig. S12 – Expression analysis of transposable elements located in the core genome or in LS regions. The box plot summarizes expression levels approximated by RNA-seq read counts, which is expressed as $\log_{10}$(RPKM +1) of genes (grey) and transposable/repetitive elements (white) for the whole genome, the core-genome and the variable genome (LS regions). Statistical differences were assessed using a Wilcoxon rank sum test.



Supplemental Fig. S13 –Dynamics of LTR transposable elements in the genome of *Verticillium dahliae* strain JR2. The divergence time of transposable elements belonging to the LTR class identified in the genome of *V. dahliae* strain JR2 (Faino et al. 2015) was estimated using the Jukes-Cantor distance calculated using the similarity between 5' and 3' Long Terminal Repeats from the same elements. The distributions of divergence times between transposable elements located in the core genome (red) and in the LS regions (blue) differ.



Supplemental Fig. S14 – Phylogenetic tree of fungal species that are related to *Verticillium dahliae*. (A) Phylogenetic relationship between 13 fungal species (FPS: *Fusarium pseudograminearum*; FFU: *Fusarium fujikuroi*; FOX: *Fusarium oxysporum* f. sp. *lycopersici*; CHI: *Colletotrichum higginsianum*; COR: *Colletotrichum orbiculare*; CGO: *Colletotrichum gloeosporioides*; ACRA: *Acremonium alcalophilum*; SOD: *Sodiomyces alkalinus*; MUCL: *Verticillium tricorpus* MUCL9792; VAA: *Verticillium albo-atrum*; MS10: *Verticillium alfalfa* MS102 LS17: *Verticillium dahliae* strain VdLs17; JR2: *Verticillium dahliae* strain JR2) was inferred using maximum-likelihood phylogeny based on a concatenated pseudo-molecule containing ~3,400 1:1:1 orthologous protein sequences identified by OrthoMCL analysis. (B) Ks distribution of orthologous gene pairs. Ks values were calculated based on genes of *V. dahliae* strain JR2 and their ortholog in each of the other species.