Supplementary Figures

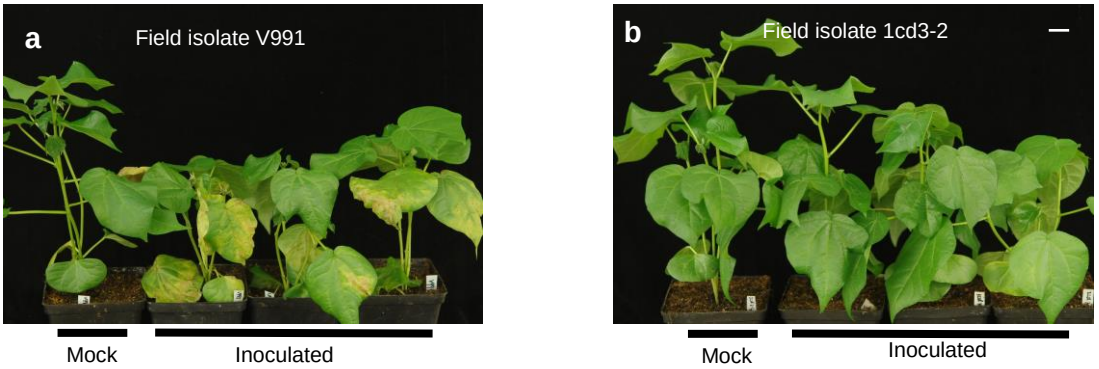


Fig. S1. The pathogenicity comparison of two reference field isolates. a The phenotype of cotton (*Gossypium barbadense*) seedlings when inoculated by V991, regarded as defoliating isolate. Mock means cotton seedlings were inoculated by water. **b** The phenotype of cotton seedlings when inoculated by 1cd3-2, regarded as non-defoliating isolate. The scale bar stands for 2 cm.

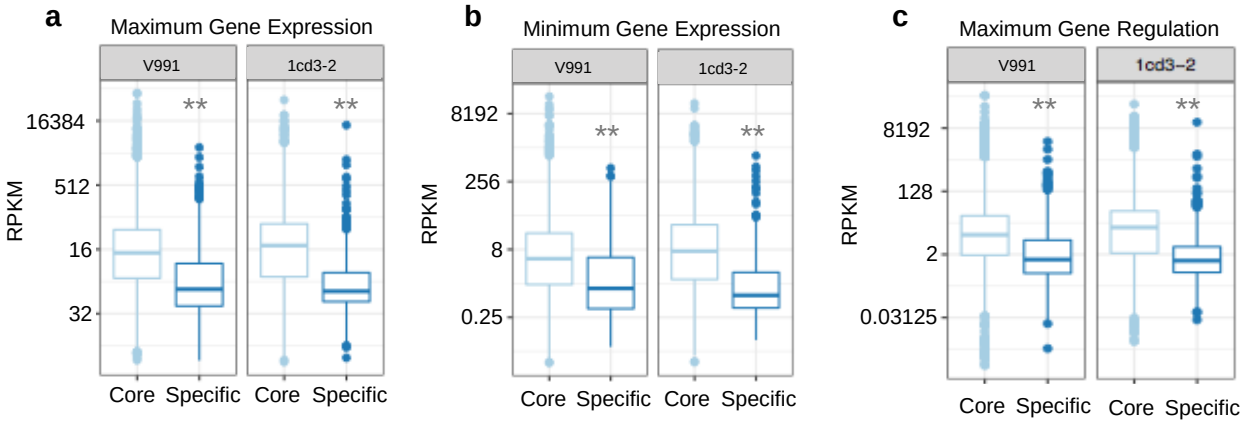


Fig. S2. specific genes have lower transcriptional levels. a Specific gene transcripts more lower during the whole infection stage with significant *P*-value < 2.2e-16 (Wilcox.test). **b** Specific gene have higher Minimum transcription level with significant *P*-value < 2.2e-16 (Wilcox.test). **c** Specific gene's expressional changing level (the difference between Maximum and Minimum expression) is higher with significant *P*-value < 2.2e-16 (Wilcox.test).

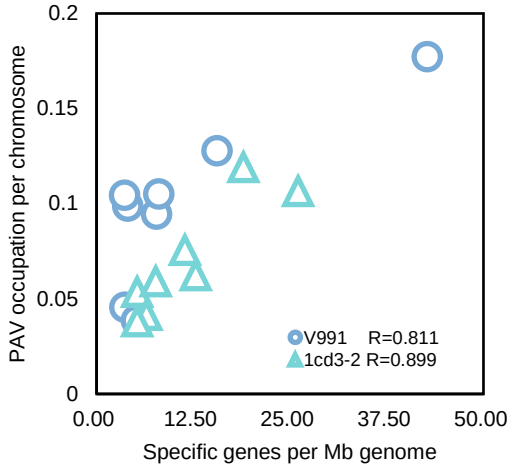


Fig. S3. High correlation of specific gene location and PAV location

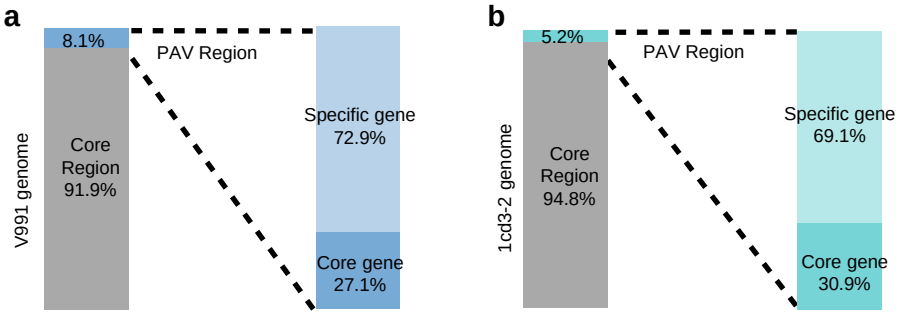


Fig. S4. Small PAV regions contain large amount of specific genes through whole genomes.

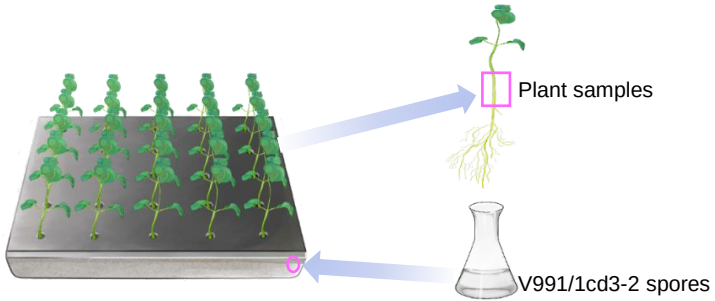


Fig. S5. The sampling mode diagram of bidirectional transcriptome on hypocotyls of cotton during the whole period of interaction (0, 3, 6, 9, 12, 15, 18, 21 days post-infection (dpi)). Total RNA was extracted from spore or infected cotton hypocotyls at 0,3,6,9,12,15,18 and dpi. The spores of the two pathogens cultured in liquid medium (Czapek-Dox) for 3 days and the hypocotyls without *V. dahliae* were used as the control. All samples were set with three independent biological replicates.

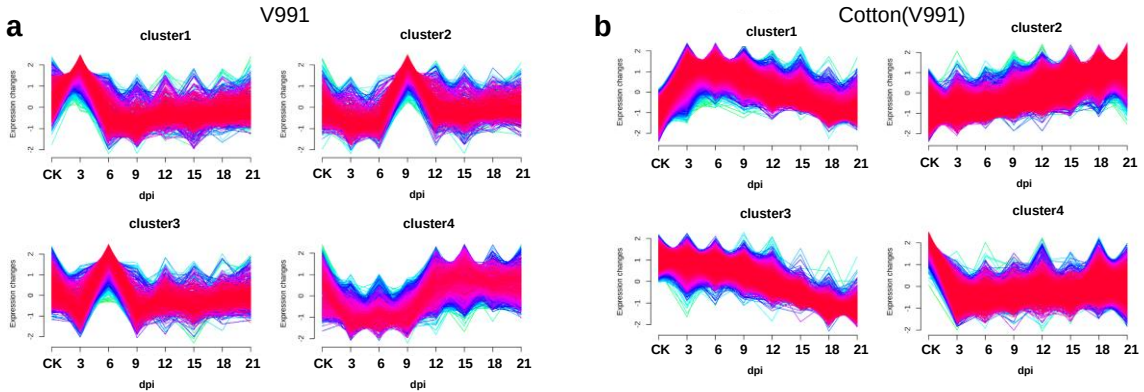
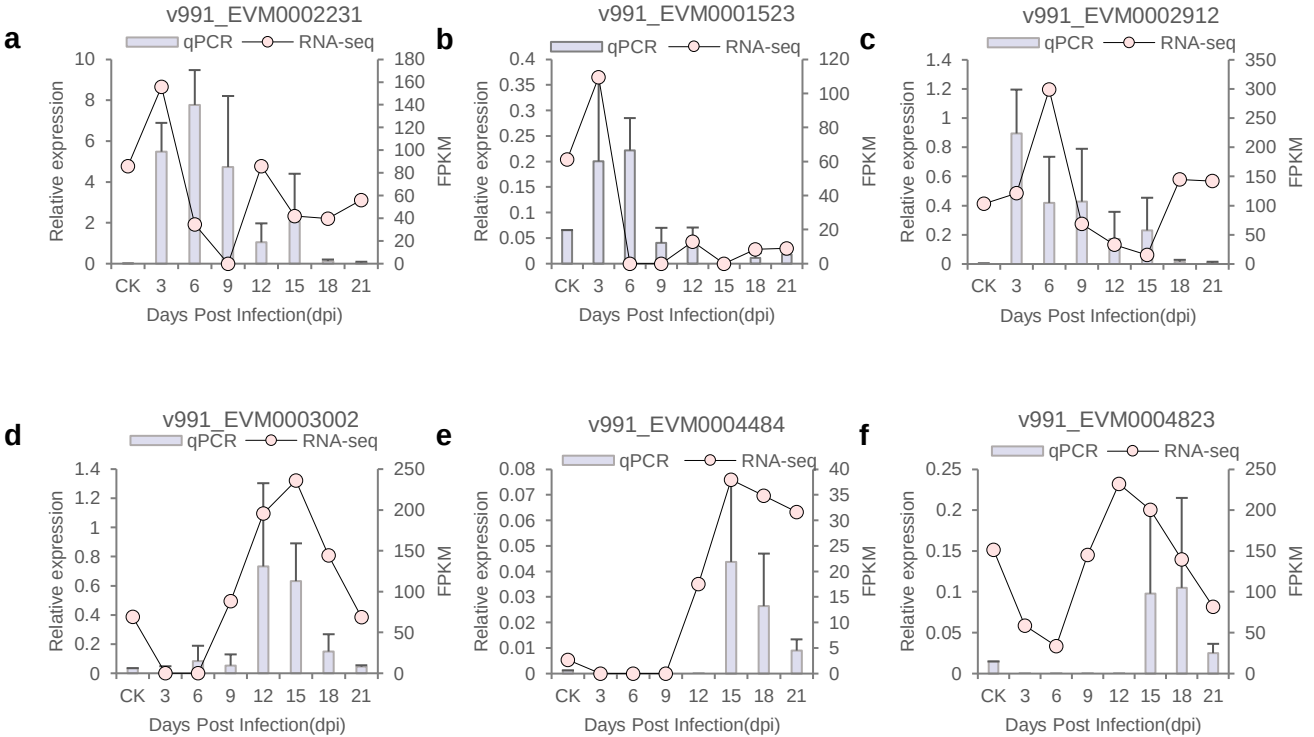


Fig. S6. Identification of classes of differentially expressed genes in transcriptomes. a Applying the fuzzy c-means algorithm to cluster 2795 differentially expressed genes in V991 transcriptomes. Clusters 1, 2, and 3 represent genes that are upregulated in stage I, clusters 4 represent genes that are upregulated in stage II. **b** Applying the fuzzy c-means algorithm to cluster 14940 differentially expressed genes in cotton (V991) transcriptomes. Clusters 1 represent genes that are upregulated in stage I, clusters 2 represent genes that are upregulated in stage II. Fuzzy c-means clustering identified 4 distinct temporal patterns of RNA expression. The x axis represents 8 time points, while the y axis represents log2-transformed, normalized intensity ratios in each time point.


a
v991_EVM0002231
qPCR
RNA-seq
Relative expression
FPKM
CK 3 6 9 12 15 18 21
Days Post Infection(dpi)
b
v991_EVM0001523
qPCR
RNA-seq
Relative expression
FPKM
CK 3 6 9 12 15 18 21
Days Post Infection(dpi)
c
v991_EVM0002912
qPCR
RNA-seq
Relative expression
FPKM
CK 3 6 9 12 15 18 21
Days Post Infection(dpi)
d
v991_EVM0003002
qPCR
RNA-seq
Relative expression
FPKM
CK 3 6 9 12 15 18 21
Days Post Infection(dpi)
e
v991_EVM0004484
qPCR
RNA-seq
Relative expression
FPKM
CK 3 6 9 12 15 18 21
Days Post Infection(dpi)
f
v991_EVM0004823
qPCR
RNA-seq
Relative expression
FPKM
CK 3 6 9 12 15 18 21
Days Post Infection(dpi)

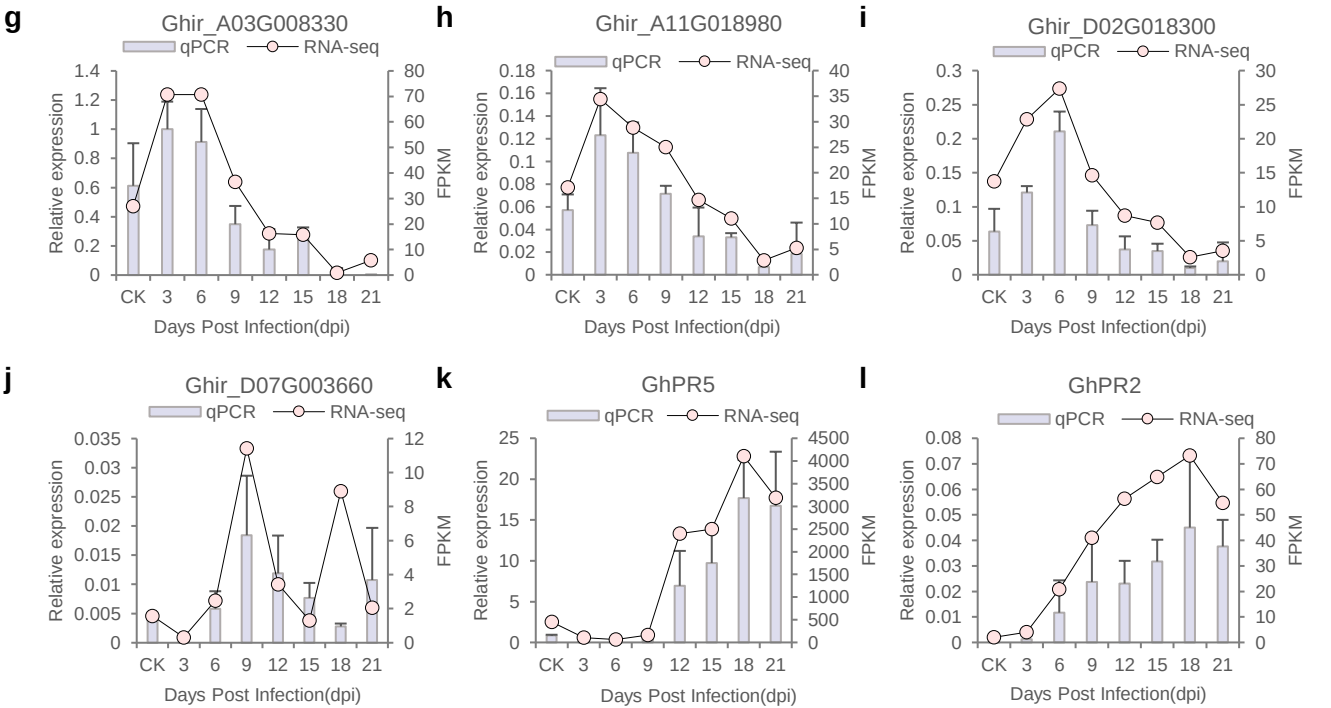


Fig. S7. Validation of gene expression patterns by qRT-PCR. a-c The relative expression level of small molecule metabolic related genes in V991 transcriptome. **d-f** The relative expression level of developmental related genes in V991 transcriptome. *v991_EVM0005718* was used as an internal reference of *V.dahliae.* **g-i** The relative expression level of cell wall biogenesis related genes in cotton (V991) transcriptome; **j-l** The relative expression level of defense response related genes in cotton (V991) transcriptome. *GhUB7* was used as an internal reference of cotton. X-axis: Days Post Infection, the y-axes indicate relative expression values (left) and FPKM values (right). The values are means ± SD, n = 3.

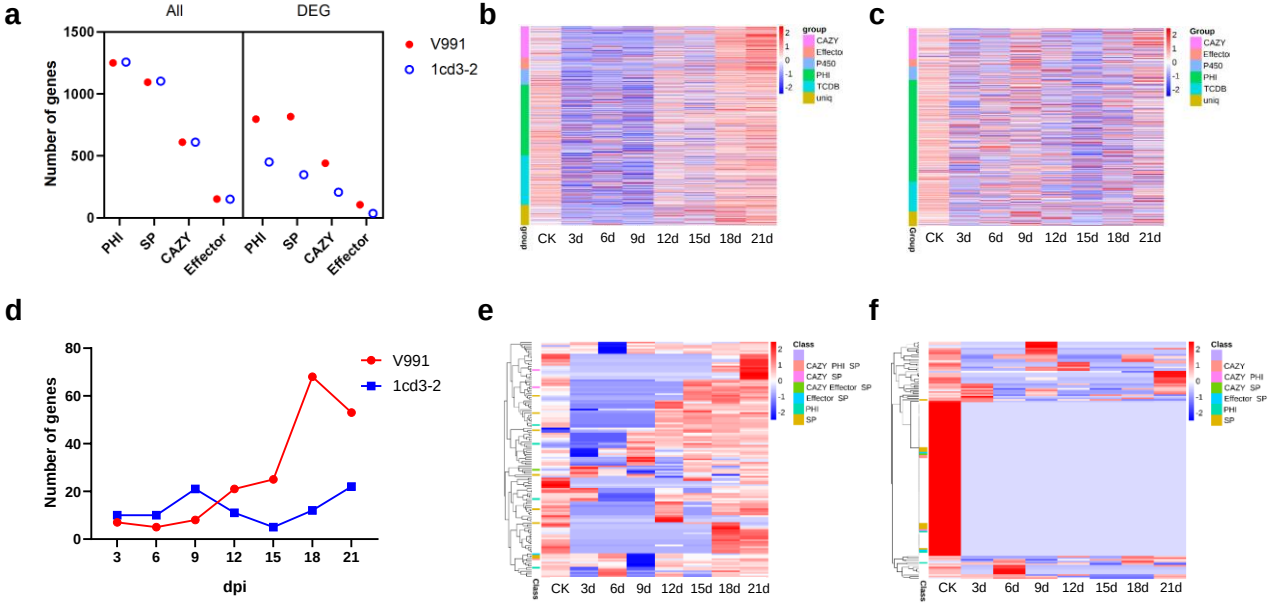


Fig. S8. Gene number statistics and gene expression analysis in V991 and 1cd3-2 transcriptomes, respectively. a Statistics of genes of different categories at the genome level (left) and transcription level (right) in V991 and 1cd3-2, respectively. **b, c** Expression profiling of pathogenicity-related genes in V991 and 1cd3-2 transcriptomes, respectively. **d** The number of difference expression genes of V991 and 1cd3-2 in the PAV regions at different time points, respectively. **e, f** Expression profiling of PAV region genes in V991 and 1cd3-2 transcriptomes, respectively.

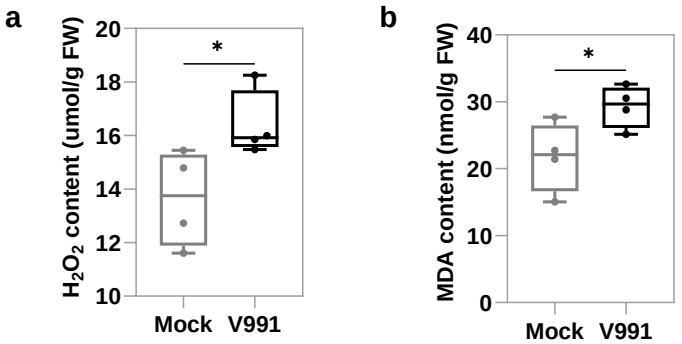


Fig. S9. H2O2 and MDA contents were significantly higher in leaves inoculated with V991. The values are means ± SD, n = 4. Statistical analyses were performed using a Student's t test: *P<0.05. All of the experiments were repeated at least three times with similar results.

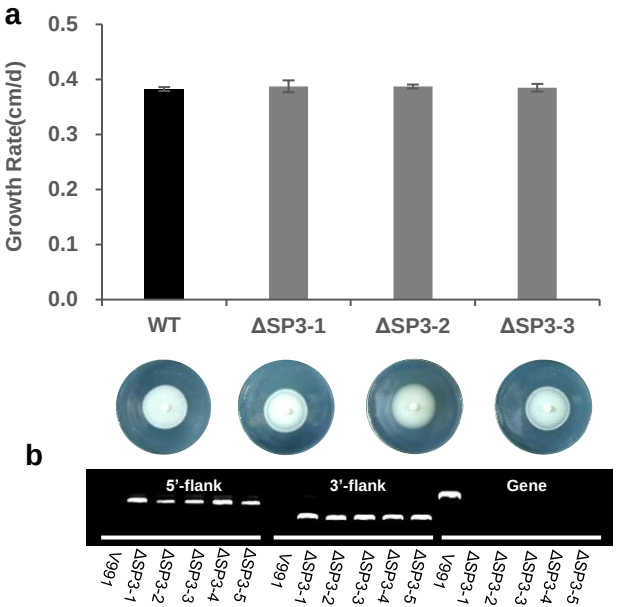


Fig. S10. Knockout of SP3 by targeted gene replacement. a Radial growth of V991 (WT) and knockout mutants (ΔSP3) on PDA agar plates. Representative cultures at 5 days post-incubation were shown. **b** PCR analysis of wild type strain V991 and knockout mutants. The genomic DNA of each strain was used to verify the 5' and 3' homologous and presence of targeted gene.

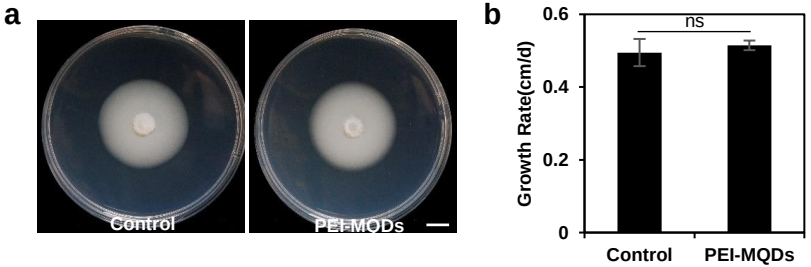


Fig. S11. Identification of growth rate of V991 on PDA agar plates without (control) and with PEI-MQDs treatment, respectively. Representative cultures at 5 days post-incubation were shown. The scale bar stands for 1 cm. The values are means ± SD, n = 3. Statistical analyses were performed using a Student's t test: ns, Not significant. All of the experiments were repeated at least three times with similar results.

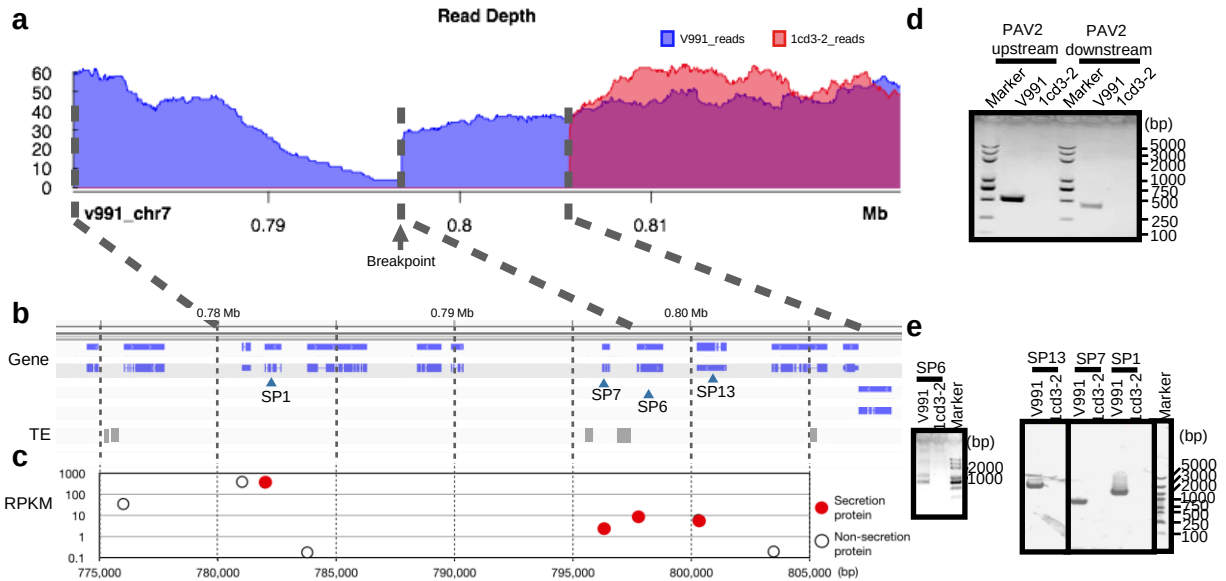


Fig. S12. PAV incidents contribute to the formation of specific gene. a-e On chromosome 7, two adjacent insertion regions totally contain 10 genes including 8 specific genes, and 4 of them are secretion proteins. **a** PAV region was verified by PacBio sequence reads separately mapping to two genomes. **b** Gene and TE distribution in the PAV region. **c** Maximum gene regulation level for each gene from upper panel. PCR validation of PAV regions (**d**) and specific genes (SP1, SP6, SP7, SP13) (**e**).

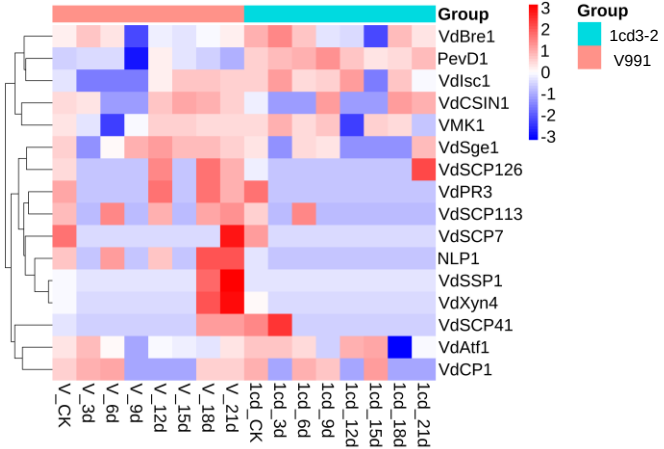


Fig. S13. Cluster heat map of reported virulence related genes in V991 and 1cd3-2 transcriptomes , respectively

Supplemental Tables

Table S1. Summary of genome assembly and annotation for *V. dahliae*

Genomic feature	V991	1cd3-2
Total scaffold (Mb)	35.7	34.5
Total contig (Mb)	35.7	34.5
Number of scaffold	13	19
Number of contigs	13	19
Largest contig (Mb)	7.8	6.7
Scaffold N50* (Mb)	3.8	3.5
Contig N50* (Mb)	3.8	3.5
Total gap (Kb)	0	0
GC content (%)	53.8	53.8
Total size of TE (Mb)	3.4	2.6
Protein-coding genes	10941	10971
Mean gene length (bp)	1855	1852
rRNA genes	83	72
tRNA genes	267	271
ncRNAs	36	35

Note: * N50, as the shortest sequence length at 50% of the genome

Table S2. Summary of core and specific genes, secreted proteins (SPs), putative effectors

Genome	V991	1cd3-2
Core genes	10411	10411
Core SPs	1068	1071
Core Effectors	108	110
Speciic genes	530	560
Specific SPs	26	32
Specific Effectors	8	7

Table S3. The average frequency of specific gene distribution in different genomic region

core/SP/whole	genome	Core region	Whole genome	PAV region
Specific gene number	V991	309	530	221
	1cd3-2	370	560	190
Genome size (Mb)	V991	32.8	35.7	2.9
	1cd3-2	32.7	34.5	1.8
Distribution Fre	V991	9.4	14.8	76.2
	1cd3-2	11.3	16.2	105.6

Table S4. Predicted secreted proteins present in V991 but absent in 1cd3-2 by PAV

Name	GeneID	Protein length	No. of cysteines	NLS sequence	Annotation
SP1	v991_EVM0000309	144	2		Putative uncharacterized protein
SP2	v991_EVM0001062	278	5		Putative uncharacterized protein
SP3	v991_EVM0003170	372	1		Cupin domain Quercetin 2,3-dioxygenase
SP4	v991_EVM0003245	398	8		Predicted protein
SP5	v991_EVM0003848	705	11		NACHT domain Vegetative incompatibility protein HET-E-1
SP6	v991_EVM0005110	358	3	PPDPKRRIEVI	GDSL-like Lipase/Acylhydrolase family
SP7	v991_EVM0005218	85	2		Putative uncharacterized protein
SP8	v991_EVM0005425	207	0		Endo-1,4-beta-xylanase activity
SP9	v991_EVM0005614	1071	17		Protein of unknown function
SP10	v991_EVM0006631	107	2		Putative uncharacterized protein
SP11	v991_EVM0009178	474	11		Predicted protein
SP12	v991_EVM0009773	389	6		Pectate lyase
SP13	v991_EVM0010514	375	1		Lyase activity
SP14	v991_EVM0010728	107	2		Putative uncharacterized protein
SP15	v991_EVM0006135	163	3		Putative uncharacterized protein

Signal peptide was predicted by SignalP (http://www.cbs.dtu.dk/services/SignalP/).
Nuclear localization signal (NLS) sequence was predicted by cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi)

Table S5. Primers used in the study

Primer name	Primer sequence 5' to 3'	Function
PAV1_upstream_F	ACTTACGGCTTCAATACGGGTTT	PCR validation of PAV1 region
PAV1_upstream_R	AGGGCGGCTACGAATCGA	PCR validation of PAV1 region
PAV1_downstream_F	ATCCTAGCGGCTATAAATAGTAAGTCCT	PCR validation of PAV1 region
PAV1_downstream_R	CAGCAACCCTAACCCTCTAACAGT	PCR validation of PAV1 region
PAV2_upstream_F	TGCGACACTCCCGACAAGC	PCR validation of PAV2 region
PAV2_upstream_R	CAAGTCTCCACCGAACAACCG	PCR validation of PAV2 region
PAV2_downstream_F	ACAACCCAGCCGTCTATGAACAC	PCR validation of PAV2 region
PAV2_downstream_R	TAGCCTGCCCAGTCAAGCCA	PCR validation of PAV2 region
SP3_F	ATGAAGCTGGCTACTACCCTC	Cloning SP3
SP3_R	GCATATTGGGGATATGTGAC	Cloning SP3
SP5_F	CGTCGGCATTTCCATAACAGTG	Cloning SP5
SP5_R	TGGCGGGAGCGAAACGT	Cloning SP5
SP8_F	GTGACTCTTGTACCTGGAAAATAGATG	Cloning SP8
SP8_R	GCGGTCAATCGTTAGGTATACATCT	Cloning SP8
SP1-F	TAGGCAAGGGAGTTGGCAG	Cloning SP1
SP1-R	AATCGTCGGTCTTTGTAGGTGA	Cloning SP1
SP6-F	GCGGGGAGTCTGGAGGTATT	Cloning SP6
SP6-R	CGAGTATGTGGTTCGCATTTTC	Cloning SP6
SP7-F	CCCTCTTCTTCCTCACCA	Cloning SP7
SP7-R	TCCCTCTAGCCTGCCTC	Cloning SP7
SP13-F	CATGTAGACAGCTTAGTTAGGGATCA	Cloning SP13
SP13-R	CCCTCAAGTAGACTCCAAGCGT	Cloning SP13
SP3_qRT_F	ACTACCCTCGGAGTATTGGC	qRT-PCR
SP3_qRT_R	CGGGATTCGCATTACTTGA	qRT-PCR
Ghir_A03G008330_qRT_F	CTCCTCCTGATGCTTACCCTT	qRT-PCR
Ghir_A03G008330_qRT_R	TTCACCGATTCAACCTTCCA	qRT-PCR
Ghir_A11G018980_qRT_F	CAACAACAGCACAGCCAAAC	qRT-PCR
Ghir_A11G018980_qRT_R	GCACTCAACCTAACCCTAACTAAA	qRT-PCR

Ghir_D02G018300_qRT_F	TGAACTGGACAGCAGGACC	qRT-PCR
Ghir_D02G018300_qRT_R	CAAGGCATCATCCCATCTAC	qRT-PCR
Ghir_D07G003660_qRT_F	ACTTCCCTGAATAACTCAAACC	qRT-PCR
Ghir_D07G003660_qRT_R	AGTCAGCCATTTCTCCCTC	qRT-PCR
GhPR5_qRT_F	CCCGTCACCCTGGTAGAGTT	qRT-PCR
GhPR5_qRT_R	TGCATATTGGCAATCCCCTG	qRT-PCR
GhPR2_qRT_F	CAATTTCTCGTTCCGGCCATGC	qRT-PCR
GhPR2_qRT_R	CCTGTGTCGATCGCTGTGGAAA	qRT-PCR
GhUB7_qRT_F	GAAGGCATTCCACCTGACCAAC	qRT-PCR
GhUB7_qRT_R	CTTGACCTTCTTCTTCTTGTGCTTG	qRT-PCR
v991_EVM0003002_qRT_F	GTGATGGCGGCGAATGA	qRT-PCR
v991_EVM0003002_qRT_R	TCTCCAGACCCACAGCACTT	qRT-PCR
v991_EVM0004484_qRT_F	AACGATCAGAGCCTTCAGAATAAA	qRT-PCR
v991_EVM0004484_qRT_R	GTGGATAAATGGGAACAGTGGA	qRT-PCR
v991_EVM0004823_qRT_F	GAAGGACAACAGCGGTGC	qRT-PCR
v991_EVM0004823_qRT_R	CCTCAAAGATCAGACGGAAA	qRT-PCR
v991_EVM0001523_qRT_F	ATGGGATCTACAACCACCAACAC	qRT-PCR
v991_EVM0001523_qRT_R	GGGACTTGAACGAGCCTGAG	qRT-PCR
v991_EVM0002231_qRT_F	ATTCTCGGCGACCTGCTC	qRT-PCR
v991_EVM0002231_qRT_R	GTCAGGGAGACTGGTAAATGGA	qRT-PCR
v991_EVM0002912_qRT_F	ACGCCAATGGCAAGAAGG	qRT-PCR
v991_EVM0002912_qRT_R	GCATTCCAGGTAGCGGTGAT	qRT-PCR
v991_EVM0005718_qRT_F	CACCTCGTTCTCCGCCTCC	qRT-PCR
v991_EVM0005718_qRT_R	TCATCCAGTGCCCCACCTAC	qRT-PCR
	CTTGCTGAGGTCTTAATTAA	Generation of knockout mutant in
SP3_UP_F	TGGCTTGAGTGATAATAACGG	*V . dahliae*
	AGTGCTGAGGCATTAATTAA	Generation of knockout mutant in
SP3_UP_R	TGAACGAGCGATGGTGTCT	*V . dahliae*
	CCCGCTGAGGACTTAATTAA	Generation of knockout mutant in
SP3_DOWN_F	GATTTCGTTGCCTGACTCACT	*V . dahliae*
	CTCGCTGAGGGTTTAATTAA	Generation of knockout mutant in
SP3_DOWN_R	CCGACGGTGTATGGATTGTT	*V . dahliae*