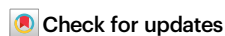


# Starship giant transposons dominate plastic genomic regions in a fungal plant pathogen and drive virulence evolution

Received: 19 February 2025

Accepted: 7 July 2025

Published online: 24 July 2025



Yukiyo Sato<sup>1</sup>, Roos Bex<sup>2</sup>, Grardiy C. M. van den Berg<sup>3</sup>,  
Parthasarathy Santhanam<sup>4</sup>, Monica Höfte<sup>2</sup>, Michael F. Seidl<sup>5</sup> &  
Bart P.H.J. Thomma<sup>1</sup>✉

*Starships* form a recently discovered superfamily of giant transposons in Pezizomycotina fungi, implicated in mediating horizontal transfer of diverse cargo genes between fungal genomes. Their elusive nature has long obscured their significance, and their impact on genome evolution remains poorly understood. Here, we reveal a surprising abundance and diversity of *Starships* in the phytopathogenic fungus *Verticillium dahliae*. Remarkably, *Starships* dominate the plastic genomic compartments involved in host colonization, carry multiple virulence-associated genes, and exhibit genetic and epigenetic characteristics associated with adaptive genome evolution. Phylogenetic analyses suggest extensive horizontal transfer of *Starships* between *Verticillium* species and, strikingly, from distantly related *Fusarium* fungi. Finally, homology searches and phylogenetic analyses suggest that a *Starship* contributed to de novo virulence gene formation. Our findings illuminate the profound influence of *Starship* dynamics on fungal genome evolution and the development of virulence.

Transposable elements (TEs, transposons) are ubiquitous mobile genetic elements in all life forms. Originally, these have been seen as selfish elements carrying only genetic information for their proliferation, but presently are appreciated to shape genome structure and function, and drive evolutionary innovations<sup>1</sup>. Whereas most TE superfamilies have simple structures and encode few proteins<sup>2</sup>, giant TEs are tens to hundreds of kilobases (kb) in size and carry tens to hundreds of cargo genes<sup>3–5</sup>.

*Starships* are giant TEs (15–700 kb) that were recently discovered in Pezizomycotina fungi, the largest subdivision of Ascomycota, and typically contain a tyrosine recombinase (YR) “captain” gene as the first gene, required for transposition, while cargo genes are variable and functionally diverse<sup>3–8</sup>. How *Starships* impact global genome evolution, and the range and extent to which *Starships* have shaped genomes

over time, remains enigmatic. *Starship* detection is technically challenging owing to the large diversity in cargo that can include abundant repeats<sup>7</sup>, and relies on the detection of the presence/absence of YR-containing inserts in orthologous sites among highly contiguous genome assemblies<sup>5</sup>. While 143 *Starships* were identified in this manner in a systematic search of 2899 fungal genomes, 10,628 “orphan” captain-like YR genes remained, suggesting many overlooked *Starships*<sup>6</sup>. *Starships* occupy up to 2.4% of the genome of the human pathogen *Aspergillus fumigatus*, show extensive presence/absence variation, and contain many differentially expressed cargo genes upon infection<sup>9</sup>. Moreover, in the plant pathogen *Macrophomina phaseolina*, 30% of chromosomal translocations, inversions, and putative chromosomal fusions occur near *Starship* insertions<sup>7</sup>. Interestingly, *Starships* can transfer horizontally between closely-related fungi of the

<sup>1</sup>Institute for Plant Sciences, Department of Biology, University of Cologne, Cologne, Germany. <sup>2</sup>Laboratory of Phytopathology, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Ghent 9000, Belgium. <sup>3</sup>Laboratory of Phytopathology, Wageningen University and Research, Droevendaalsesteeg 1, 6708PB Wageningen, the Netherlands. <sup>4</sup>Agriculture and Agri-Food Canada, Morden Research and Development Centre, Morden, MB, Canada. <sup>5</sup>Theoretical Biology & Bioinformatics Group, Department of Biology, Utrecht University, Utrecht, the Netherlands. ✉e-mail: [bthomma@uni-koeln.de](mailto:bthomma@uni-koeln.de)

same order, and can transfer important traits such as pathogenicity<sup>8,10–16</sup>.

Pathogens and their hosts typically engage in molecular arms races, with the pathogen exploiting secreted virulence factors (effectors) to mediate host colonization, while hosts employ immune receptors for pathogen interception<sup>17–19</sup>. To avoid recognition, pathogen effector catalogs are highly dynamic and variable, mediated by a “two-speed genome” organization in which virulence genes co-localize in highly plastic genomic regions that are enriched in repetitive elements and particular epigenetic features<sup>20–26</sup>. Accordingly, the fungus *Verticillium dahliae* that causes vascular wilt disease in hundreds of hosts<sup>27–29</sup> contains plastic ‘adaptive genomic regions’ (AGRs)<sup>30</sup> that are enriched in virulence genes<sup>31–36</sup>, transcriptionally active TEs<sup>30,37,38</sup>, and structural variations<sup>32,37–39</sup>, associated with a unique chromatin profile and physical interactions in the nucleus<sup>30,40–42</sup>. Besides *V. dahliae*, the Pezizomycotina *Verticillium* genus contains nine additional plant-associated species<sup>28</sup>. Thus far, two *Starships* have been identified in a single strain of *V. dahliae*<sup>6</sup>. Here, we queried 56 highly contiguous *Verticillium* genome assemblies for *Starships* to analyze their association with the evolution of AGRs and virulence on plant hosts.

## Results

### A wealth of *Starships* occurs in the *Verticillium* genus

To identify *Starships* across the *Verticillium* genus, we collected 56 high-quality *Verticillium* genome assemblies, comprising 36 *V. dahliae* strains and 20 strains of the nine remaining species (Fig. 1a and Supplementary Data 1, 2) and queried these genomes for *Starships* using “Starfish”<sup>6</sup>. We uncovered 54 *Starships* that belong to 24 haplotypes of 14 naves of seven families (Fig. 1a, b and Supplementary Data 3). Between one and three *Starships* were detected in 33 of the 56 strains belonging to seven of the ten *Verticillium* species (Fig. 1a, b). Thus, most *Verticillium* genomes contain a *Starship*, and several genomes even contain multiple. Moreover, as these *Starships* range from 17 to 625 kilobases (kb) (Fig. 1c), and larger ones typically contain multiple YR genes of different naves (Fig. 1c, d), these likely represent nested *Starship* insertions. Thus, the final number of *Verticillium* *Starships* is likely under-estimated.

*Verticillium* genomes exhibit extensive large-scale genomic rearrangements that may have affected the integrity of prior inserted *Starships*<sup>32,37,43</sup>. However, Starfish cannot identify fragmented *Starships*, rearranged *Starship* insertion sites, or *Starship* insertions into lineage-specific regions<sup>6</sup>. To identify such *Starships*, we queried for reference captain and captain-like YRs previously identified in Pezizomycotina genomes<sup>6</sup>, revealing 2–18 homologs in each strain, amounting to 556 YR genes of 38 naves and seven families (Supplementary Fig. S1 and Supplementary Data 4). Only 21% of them belong to *Starships* identified by Starfish (Fig. 1e), while the remaining 79% could point to unidentified *Starships*. Additionally, to detect potential *Starships* and *Starship*-like regions that may have been overlooked due to genomic rearrangements and insertion into lineage-specific regions, we identified regions syntenic to *Starships* as “*Starship* regions”. Hereafter, “*Starships*” refer to those identified using Starfish, while “*Starship* regions” collectively refer to *Starships* regions plus syntenic regions. Intriguingly, such *Starship* regions occur in all strains (Supplementary Data 5), and half of the captain-like YR genes that did not occur in *Starships* appeared in such *Starship* regions (Fig. 1e). Thus, we reveal abundant *Starships* and their remnants in the *Verticillium* genus.

### *Starships* are hotspots of large-scale genomic rearrangements

To detail how genomic rearrangements affected *Starships*, we compared telomere-to-telomere genome assemblies of *V. dahliae* strains JR2 and VdLs17 that comprise dozens of large-scale genomic rearrangements<sup>32,37,44</sup>. In strain JR2, we detected three large *Starships* of 0.50–0.54 Mb each, belonging to haplotypes *Ar1h1*, *Ar4h1*, and *Se1h1*, plus additional *Starship* regions, collectively accounting for

5.3% (1.92 Mb) of the genome (36.15 Mb) (Fig. 2a). Although no *Starships* were detected in VdLs17 by Starfish, *Starship* regions account for 2.7% (0.96 Mb) of the genome (35.97 Mb) (Fig. 2a). Intriguingly, 60% of the inter-chromosomal rearrangement breakpoints between these strains occurred in *Starship* regions (Fig. 2a and Supplementary Fig. S2). Accordingly, *Starship* regions were mainly detected in AGRs that are enriched in such rearrangements<sup>32,37</sup> (Fig. 2a, b). In strain JR2, 92% (1.77 Mb) of the *Starship* regions colocalized with AGRs (Supplementary Data 6). Moreover, 53% of the total AGR complement belongs to *Starships*. In VdLs17, 94% (0.90 Mb) of the *Starship* regions colocalized with 22% of the AGR complement (Supplementary Data 6).

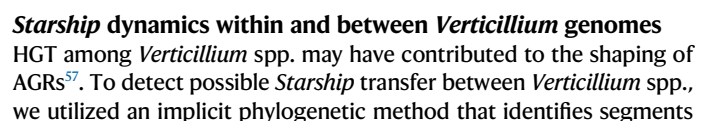
Since significant AGR proportions could not be assigned to *Starships* in JR2 (47%) and VdLs17 (78%), we determined whether *Starship* regions in AGRs display different characteristics than other AGRs. Pairwise genomic alignments of the JR2 genome with those of the other 35 *V. dahliae* strains revealed significantly lower alignment coverage in *Starship* regions (median 31%) than in other AGRs (median 85%) and core genomic regions (median 93%) (Fig. 2c and Supplementary Fig. S3), indicating enhanced presence/absence variation or sequence divergence in *Starship* regions. Moreover, we revealed an extreme enrichment in structural variation, particularly concerning translocations, inversions, and duplications, in *Starships* (Fig. 2d). Regions lacking alignment accompanying these rearrangements were significantly larger in *Starship* regions (median 14.7 kb) than in other AGRs (median 5.9 kb) and core regions (median 1.7 kb) (Fig. 2e and Supplementary Fig. S4a), underscoring the extensive presence/absence variation in *Starship* regions. Intriguingly, some genomic rearrangements even occurred between *Starships* as the *Ar1h1* *Starship* in *V. dahliae* strain GF1192 is largely syntenic to the JR2 *Ar1h1* *Starship*, but synteny lacks at four sites that are syntenic to parts of the JR2 *Ar4h1* *Starship* (Fig. 2f and Supplementary Fig. S4b), leading to *Starship* diversification.

### *Starships* display typical traits of plastic genomic regions

Like plastic genomic regions of many filamentous plant pathogens<sup>20–24</sup>, *V. dahliae* AGRs are enriched in effector genes, *in planta* induced genes with facultative heterochromatic histone modifications (H3K27me3)<sup>30,32,41</sup>, and repetitive elements including active TEs<sup>37,38</sup>. Intriguingly, the previously characterized effector genes *Ave1*<sup>21,33</sup> and *Av2*<sup>24,45</sup> occur in *Ar1h1* and *Ar4h1* *Starships* (Fig. 3a, b). While no difference in *in planta* gene induction and H3K27me3 levels could be observed between *Starship* regions and other AGRs (Fig. 3c, d), the average TE density and expression were higher in *Starship* regions than in other AGRs (Fig. 3e–g). TEs in many fungi are typically inactivated by repeat-induced point mutation (RIP) that introduces cytosine-to-thymine mutations during sexual stages<sup>46,47</sup>. The majority of TEs in other AGRs and core regions showed the signature of RIP (positive composite RIP index (CRI) values<sup>48</sup>), but the majority of TEs in *Starship* regions did not (Fig. 3h), consistent with their expression levels (Supplementary Fig. S5). *V. dahliae* AGRs have furthermore been characterized by segmental duplications<sup>37</sup> that physically interact in the nucleus<sup>42</sup>. Interestingly, consistent with the segmental duplication pattern (Figs. 2d, 3i), such bipartite long-range chromatin interactions occur between the *Ar1h1* and *Ar4h1* *Starships*, between nested *Starships* within the *Ar1h1* and *Ar4h1* *Starships*, and between the *Se1h1* *Starship* and syntenic *Starship* regions in the JR2 genome (Fig. 3j). Collectively, our data indicate that *Starships* in *V. dahliae* are strongly associated with traits that characterize AGRs.

### *Starships* carry virulence cargo genes

To assess whether previously characterized virulence genes besides *Ave1* and *Av2* appear as cargo, we queried the *Starships* for *Verticillium* genes that are described in the Pathogen–Host Interactions Database (PHI-base)<sup>49</sup>. This revealed that the G-LSR2 region, proposed to confer





**Fig. 1 | Diverse *Starships* populate the *Verticillium* genus.** **a** The tree in black shows the phylogeny of the 56 strains used in this study across the *Verticillium* genus, divided into the Flavexudans (FE) and Flavonoxudans (FNE) clades<sup>28</sup>, based on whole-genome sequence alignments. Circle colors indicate the ten *Verticillium* spp. while the label comprises species abbreviation followed by strain name. The tree in gray shows only the *V. dahliae* strains at increased resolution. Scale bars indicate phylogenetic distances expressed as nucleotide substitutions per site. **b** Repertoires of *Starship* haplotypes (hap.) per strain. *Starships* were classified according to previous studies<sup>5,6</sup>. Columns indicate *Starship* haplotypes defined by *k*-mer similarity and named according to captain navis and family (Supplementary Fig. S1b), whereas heatmap colors show haplotype member counts. “Id” refers to identical *Starships* (coverage and nucleotide sequence identity >98%) within a haplotype. **c** Size of the different *Verticillium* *Starships*. Points indicate individual

*Starships* listed in Supplementary Data 3. Gray crossbars and error bars indicate the median and 95% confidence interval range of *Starship* lengths for each navis. **d** Captain/captain-like tyrosine recombinase (YR) navis repertoires in *Starships* with multiple YR genes. **e** Captain/captain-like YR gene classification per strain where “*Starship* (captain)” indicates YR genes located as the first gene at the 5'-terminus of a *Starship* for which both borders could reliably be identified with the Starfish tool<sup>6</sup>, “*Starship* (non-captain)” refers to YR genes located at other sites in a *Starship*, suggesting nested *Starships* with unidentified boundaries. Furthermore, “other *Starship* region” indicates YR gene presence in regions that could not reliably be identified as *Starship* with Starfish, but that are syntenic to reference *Starships*. Finally, “unaffiliated” refers to YR genes that cannot reliably be affiliated with a *Starship* region.

with increased sequence similarity over the genome-wide average nucleotide identity (ANI)<sup>58</sup>. This analysis revealed that several *Starships*, such as those belonging to haplotypes *Se1h1* and *Ar1h1*, as well as syntenic *Starship* regions have higher ANI levels than the genome-wide average between the *Verticillium* species in which they occur (Fig. 4a, b and Supplementary Data 8). The most conspicuous is the *Ar1h1* *Starship* that carries *Ave1*. The *Ar1h1* *Starship* and its syntenic regions share over 99% identity over 98% coverage across 0.5 Mb between *V. dahliae* and *V. nonalfalfae*, while the genome-wide average identity between these species is only 93% (Fig. 4c and Supplementary Data 8).

To further support *Starship* mobility, we compared synteny of orthologous *Starship* flanking regions, revealing that *Se1h1* and *Ar1h1* *Starships* occur in several non-homologous genomic regions in different *Verticillium* clades (Fig. 4c, d and Supplementary Fig. S7). For instance, the *Ar1h1* *Starship* and its syntenic regions that occur in nine *V. dahliae* strains and in two *V. nonalfalfae* strains collectively inserted into regions that are orthologous to five different JR2 chromosomes (Fig. 4d), suggesting at least four *Ar1h1* *Starship* movements. *Se1h1* *Starships* in three *V. dahliae* strains, one *V. longisporum* strain, and one *V. nonalfalfae* strain collectively occur in regions that are orthologous to four different JR2 chromosomes (Supplementary Fig. S7c). *Se1h1* *Starships* in some strains appear to have been generated by rearrangements between regions orthologous to two JR2 chromosomes (chr2 and chr5) but have been inserted into two other different chromosomal regions (JR2 chr1 and chr4), suggesting at least two movements (Supplementary Fig. S7c).

Comparisons among orthologous *Starship* regions across *Verticillium* strains revealed various diversification patterns (Supplementary Fig. S8). Besides nested *Starship* insertions (e.g., *Ar3h1* *Starship* in *Ar1h1* *Starship*, Supplementary Fig. S8a and Fig. 1c, d), we typically observed gain and loss of cargo elements in TE-enriched *Starships* (e.g., *Pr1h1* *Starships*, Supplementary Fig. S8b, c). We furthermore observed invasion of orthologous sites by different *Starships* that contain orthologous captains, but otherwise lack synteny (e.g., *Ar1h1* and *Ar1h2* *Starships*, Supplementary Fig. S8d), which points towards independent insertions as orthologous captains target particular sequences as *Starship* insertion sites<sup>6,8,59</sup>.

### Cross-order horizontal *Starship* transfer

To detect possible horizontal *Starship* transfer between *Verticillium* and other fungi, we queried all Pezizomycotina genomes available in the GenBank WGS database (10,113 genomes from 668 genera, Supplementary Fig. S9a and Supplementary Data 9) with *Verticillium* *Starships*. Intriguingly, regions syntenic to *Ph1h1* and *Ph1h2* *Starships* that were detected in only three *V. dahliae* strains (CQ2, VD991, and XJ592) (Fig. 4a, b), were found in the genomes of various *Fusarium* species, some of which contain regions syntenic to the G-LSR2 region (*Ph1h1* *Starship* cargo) that was associated with *V. dahliae* virulence<sup>50</sup> (Fig. 5a and Supplementary Fig. S9b). Since *Fusarium* belongs to Hypocreales and synteny to these *Verticillium* *Starships* lacks in non-*Verticillium* genomes of the Glomerellales to which *Verticillium* spp.

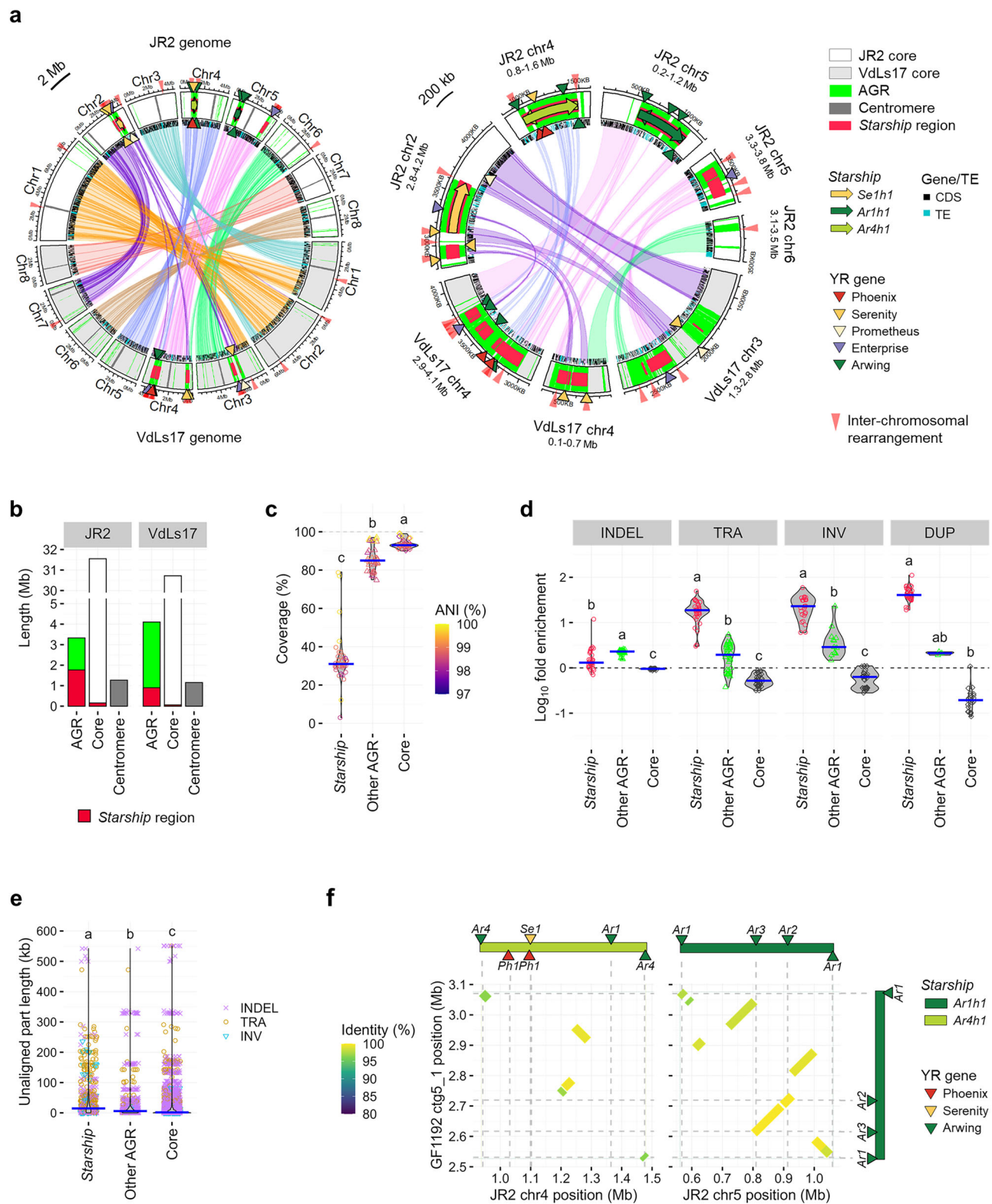
belong, these results suggest horizontal transfer of *Ph1h1* and *Ph1h2* *Starships* between *Verticillium* and *Fusarium*.

To identify the directionality of horizontal transfer, we analyzed a *k*-mer-based phylogeny of *Starships*<sup>60</sup>. To this end, we identified 52 *Starships* in 283 high-quality *Fusarium* genome assemblies (Supplementary Data 10, 11) and analyzed their phylogeny together with the 54 *Verticillium* *Starships*, revealing that most *Starships* clustered according to genus (Fig. 5b). In contrast, the *Verticillium* *Ph1h1* and *Ph1h2* *Starships* phylogenetically localized in the *Fusarium* *Starship* clade (Fig. 5b), suggesting *Fusarium* to *Verticillium* transfer. Moreover, we identified a near-identical *Ph1h2* *Starship* that shares over 99% nucleotide identity over 50 kb in three *V. dahliae* strains and in *F. keratoplasticum* (Fig. 5c). The *Starship* flanking regions are syntenic in three phylogenetically clustered *V. dahliae* strains, suggesting that the *Ph1h2* *Starship* invaded *V. dahliae* from *F. keratoplasticum* before these strains diverged (Fig. 5c).

To further explore potential *Starship*-associated HGTs overlooked by the synteny search, we tested HGT signatures for each *Starship* cargo element using the Alien index (AI) that detects higher similarity to outgroup than ingroup orthologs<sup>61</sup>. We queried orthologs of *Verticillium* genes in the Pezizomycotina genomes and calculated AI scores with non-*Verticillium* Glomerellales genera as ingroup and the other 74 orders as outgroup. This showed that the median AI for JR2 genes is higher in *Starship* regions (45) than in other AGR (0) and core regions (−59), and that the proportion of genes with AI ≥45 (indicative of HGT) is higher in *Starship* regions (19%) and other AGRs (17%) than in core regions (4%), suggesting cross-order HGT in *Starship* regions and other AGRs (Fig. 5d). Among the virulence-associated cargo genes, *Av2* as well as the G-LSR2 genes showed AI >45 with hits in the various *Fusarium* spp. of the Hypocreales while lacking ingroup hits (Fig. 5e). Horizontal *Av2* transfer was further supported by phylogenetic analyses in which *Verticillium* orthologs were detected in a clade within a larger clade of *Fusarium* orthologs (Supplementary Fig. S10a, b). *Fusarium* *Av2* orthologs were found proximal to captain-like YR genes, which suggests *Starship* association (Supplementary Fig. S10c). We also tested the HGT signature of transcriptionally active *Verticillium* TEs<sup>62</sup> that occur in *Starship* regions (Supplementary Fig. S11a). Surprisingly, seven out of nine cargo TEs showed AI >45 with the best outgroup hits in five orders, including Hypocreales, while six of them lacked ingroup hits (Fig. 5e). Orthologs of the five cargo TEs also occurred in *Fusarium* *Starships* (Supplementary Fig. S11b), suggesting horizontal transfer via *Starships*. Collectively, our results provide evidence for horizontal transfer of various *Starship* cargo elements between *Verticillium* and fungi that even belong to other orders.

### *Starships* mediate de novo gene birth

*Starships* carry diverse lineage-specific genes<sup>7</sup>, yet their origin remains enigmatic. We pursued the evolution of the *NLP6* effector gene that occurs in a limited number of *V. dahliae* strains, including VdLs17<sup>63</sup>, and was detected near a *Starship* region (Supplementary Fig. S12a). *NLP6* belongs to the family of necrosis- and ethylene-inducing peptide



1 (Nep1)-like proteins (NLPs), many members of which confer virulence<sup>63,64</sup>. As multiple NLP paralogs occur in *Verticillium* genomes<sup>63</sup>, we first analyzed the phylogeny of NLP homologs in the 56 *Verticillium* genomes. NLP6 occurs in the same clade as NLP3 orthologs which were detected in all 56 strains (Fig. 6a). Intriguingly, NLP6 is more closely related to NLP3 orthologs of *V. nubilum* and species of the FE clade than to those of *V. dahliae* or other species of the FNE clade (Fig. 6a). The similarity between NLP3 orthologs and NLP6 occurs at the C- but not at the N-terminus (Supplementary Fig. S12b). Rather, the

3' end (115-450 nt) of *NLP6* shares 90% nucleotide identity with the 3' end (379-714 nt) of *V. nubilum* *NLP3*, which greatly exceeds the genome average (82%) (Fig. 6b, c, Supplementary Fig. S12c, and Supplementary Data 8), suggesting that the 3' part of *V. dahliae* *NLP6* is derived from a horizontally transferred *NLP3* ortholog. To explore the origin of the 5' part (1-114 nt) of *NLP6*, we queried the *Verticillium* genomes and detected hits with 96% nucleotide identity in non-coding regions in *Se1h1* *Starships* in four *Verticillium* species, which we termed *tNLP6* (for *truncated NLP6*) (Fig. 6d-f). Intriguingly, three *tNLP6* copies were

**Fig. 2 | *Starships* are hotspots of genomic rearrangements.** **a** Circular plots showing the locations of *Starship* regions, adaptive genomic regions (AGRs), and genomic rearrangements between *V. dahliae* strains JR2 (upper eight chromosomes) and VdLs17 (lower eight chromosomes). Tracks are filled with colors representing either core, AGR, or centromeric regions, overlaid with bold lines and arrows representing *Starship* regions and *Starships* colored by haplotype. Overlapping arrows in a single region indicate that the *Starship* orientation is not determined due to the presence of captains at the 5' end of both strands. Regular triangles point to the captain and captain-like tyrosine recombinase (YR) genes, colored by family and annotated with navis identification (ID). The overlaps among these elements and rearrangement breakpoints are shown in Supplementary Fig. S2 at a higher resolution. Colored bands at the inner edge of tracks represent genetic elements grouped into protein coding sequence (CDS) and transposable element (TE). Ribbons connect syntenic regions (>80% nucleotide sequence identity over 10 kb). **b** Total length of genomic compartments in *V. dahliae* strains JR2 and VdLs17. **c** Violin plots depicting the sequence alignment coverage determined by comparing JR2 genomic compartments against 35 *V. dahliae* genomes (Supplementary Fig. S3). Points indicate the coverage, and the color represents genome-wide average nucleotide identity (ANI) for each genome alignment ( $n = 35$ ), while blue crossbars indicate the median values. Different letter labels indicate significant

differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). **d** Violin plots depicting the fold-enrichment of structural variations (SVs) (insertion and deletion (INDEL) without size definition, translocation (TRA), inversion (INV), and duplication (DUP)) in the three genomic compartments compared with the genome-wide average. SVs were determined by comparing the genome of *V. dahliae* strain JR2 against the 35 *V. dahliae* genomes. Points indicate the fold-enrichment determined for each genome ( $n = 35$ ), while blue crossbars indicate the median values. Different letter labels indicate significant differences for each genetic variation (two-sided Dunn's test, adjusted  $p < 0.05$ ). **e** Violin plots depicting length of unaligned regions accompanied by SVs in JR2 genomic compartments. Points indicate the length of every gap ( $n = 644$  in *Starship* regions,  $n = 1163$  in other AGRs, and  $n = 8877$  in core regions) found between the JR2 genome and 35 *V. dahliae* genomes in a color and shape representing SV type, while blue crossbars indicate the median values. Different letter labels indicate significant differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). **f** *Starship* rearrangements between the genomes of *V. dahliae* strains JR2 (X-axis) and GF1192 (Y-axis). Diagonal lines indicate synteny, while the color represents nucleotide sequence identity. Bars and triangles aligned to the plots indicate the positions of *Starships* and YR genes. The left and right plots are superimposed in Supplementary Fig. S4b to make the breakpoints of inter-chromosomal rearrangements clearer.

detected around *NLP6* in VdLs17 (Fig. 6e). Collectively, these results suggest that *NLP6* could be formed through the fusion of a non-coding *Starship* element with the 3' part of an *NLP3* ortholog through a genomic rearrangement (Fig. 6g). Interestingly, although a role of *NLP3* in fungal virulence could not be demonstrated, deletion of *NLP6* results in reduced symptom development in *V. dahliae*-inoculated tomato plants (Fig. 6h and Supplementary Fig. S13). Collectively, these results demonstrate that *Starships* mediated the emergence of a virulence-associated lineage-specific effector gene in *V. dahliae*.

## Discussion

Here, we reveal the unprecedented impact of *Starships* on fungal genome evolution through four main findings. First, every *Verticillium* genome contains *Starships* or their remnants. Secondly, these *Starships* compose the vast majority of AGRs that govern pathogenicity on plant hosts. Thirdly, extensive horizontal *Starship* transfer occurred between *V. dahliae* and phylogenetically diverse fungi. Fourthly, *Starships* contributed to the de novo generation of a novel virulence gene. Thus, *Starships* are not only instrumental for *Verticillium* genome evolution, but also fundamental to evolutionary innovations that led to plant pathogenicity.

Thus far, little is known about *Starship* abundance in individual fungal genomes. We identified 24 *Starship* haplotypes in 56 *Verticillium* genomes. These *Starships* generally occur in multiple *Verticillium* strains, indicating invasion before strain diversification. While some remained intact, others got disrupted by genomic rearrangements. The abundance of *Starships* exceeds the abundance previously recorded in *A. fumigatus*, where 20 *Starships* and their remnants were identified in a survey of 519 strains, and not all strains contained *Starships*<sup>9</sup>. In *M. phaseolina*, four *Starships* and remnants were identified in 12 isolates<sup>7</sup>. Overall, our study reveals an unprecedented abundance of *Starships* across *Verticillium* genomes.

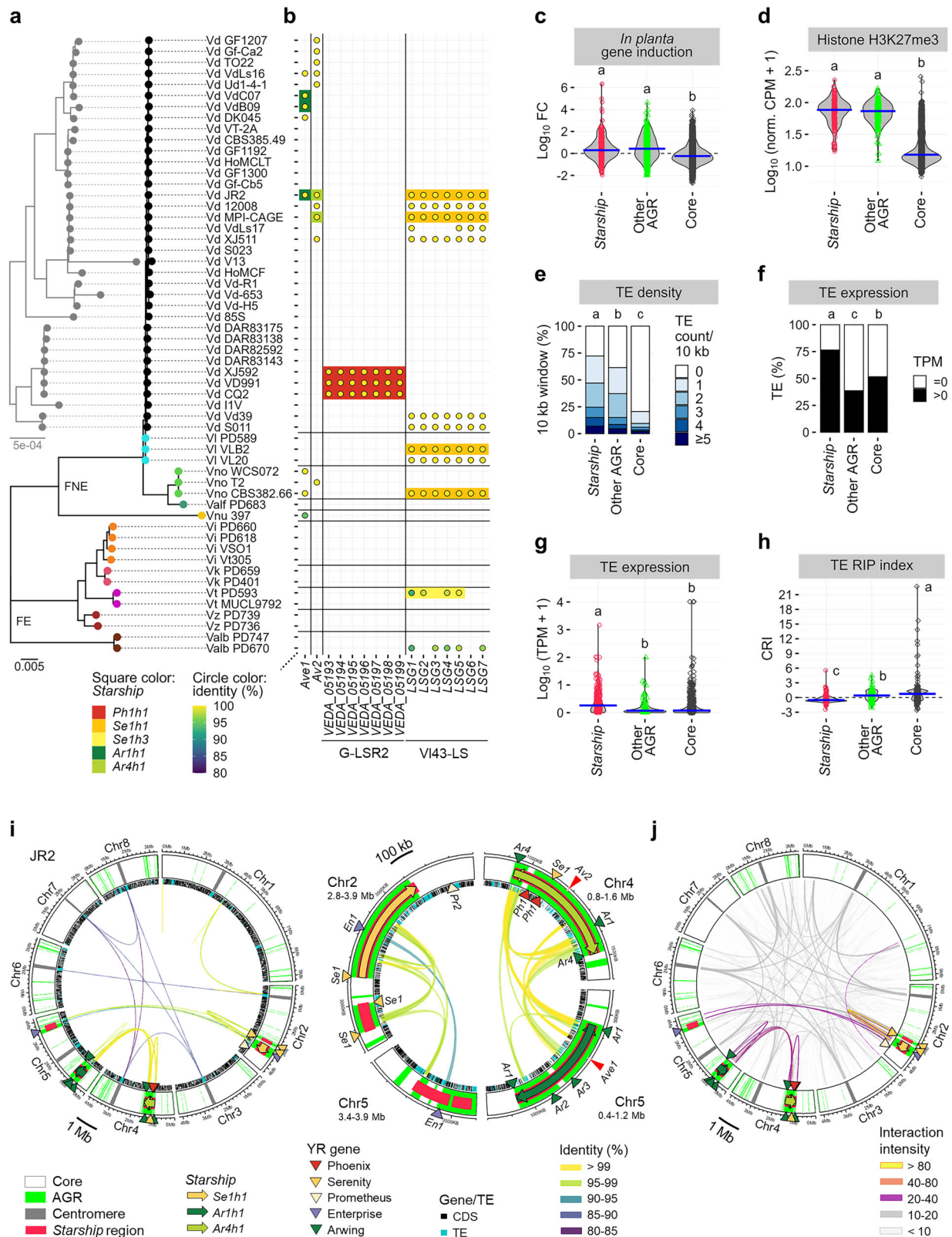
We could unequivocally demonstrate that *Starships* make up about half of AGRs in the *V. dahliae* JR2 genome, while the remaining AGRs could not be assigned to *Starships*. AGRs were initially identified as lineage-specific genomic regions by comparisons among *V. dahliae* strains<sup>32</sup>, while additional AGRs were subsequently identified based on their characteristic chromatin profile<sup>30</sup>. Interestingly, the *Starship* regions identified in this study correspond to the initially identified AGRs. However, besides their chromatin profile, like *Starships*, also the non-*Starship* AGRs are enriched in horizontally transferred genes and genomic rearrangements. Thus, we speculate that these AGRs may be derived from unidentified *Starships* that became disrupted by genomic rearrangements. In support of this hypothesis, *Verticillium* genomes

likely still contain many unidentified *Starship* regions, given 211 captain-like YR genes unassigned to known *Starship* regions in the 56 genomes, the majority (127/211) of which occur in AGRs (Supplementary Data 4). Overall, we conclude that *Starships* are the major constituents of AGRs in *V. dahliae*.

Many filamentous plant pathogens carry plastic genomic regions similar to AGRs<sup>20–24</sup>, but how these regions evolved remains elusive<sup>26</sup>. Although several *Starships* have been identified in Pezizomycotina plant pathogens<sup>7,12,14,15,65–68</sup>, the association of these *Starships* with the typical plastic genomic regions has not been noted previously. The evolution of plastic genomic regions is known to be associated with the activity of smaller TEs<sup>20,26</sup>. Our results show that TEs in *Starship* regions are more highly expressed and show weaker RIP signatures than those in the other regions of the *V. dahliae* genome. RIP mutates TEs during sexual stages in Pezizomycotina fungi<sup>47</sup>. As *V. dahliae* is presumed to reproduce asexually<sup>69</sup>, it has been hypothesized that TEs showing RIP signatures could have been mutated in its sexually active ancestor in which RIP was active<sup>30</sup>. We have previously shown that TEs in lineage-specific AGRs, which correspond to *Starship* regions identified in this study, are enriched in evolutionarily younger TEs that proliferated after the speciation of *V. dahliae*, whereas other regions are enriched in older TEs that proliferated before speciation<sup>37</sup>. Therefore, the enrichment of transcriptionally active TEs in *Starship* regions could be due to their invasion of *V. dahliae* after the inactivation of TEs in other genomic regions. In support of this hypothesis, our results suggest the horizontal transfer of active TEs in *Starships*. These findings highlight the close relationships among plastic genomic regions, *Starships*, and the activity of other TEs.

HGT has driven virulence evolution in diverse fungal pathogens<sup>70,71</sup>. HGT between phylogenetically close fungi is thought to be mediated by anastomosis and horizontal chromosome transfer, while the mechanisms of HGT between phylogenetically distant fungi remain elusive<sup>70,71</sup>. Our phylogenetic analyses suggest the horizontal *Starship* transfer between *V. dahliae* and *V. nonalfalfae* on the one hand, and distantly related *Fusarium* fungi on the other hand. In addition, *Starship* regions in *V. dahliae* are enriched in cargo elements horizontally transferred to/from fungi belonging to further diverse Pezizomycotina orders. Thus, while previous studies have shown *Starship*-mediated HGT between fungi within the same order<sup>8,10–16</sup>, our study extends this observation by revealing that *Starships* mediated HGT between phylogenetically distant fungi of different orders. There are many reports of cross-order HGT between diverse Pezizomycotina pathogens, though the mechanism remains generally unknown<sup>72–78</sup>. Further identification of *Starships* in diverse taxa may





help to understand the extent to which *Starships* have mediated such HGT events.

Pathogens exploit diverse effectors to colonize hosts, but it often remains unclear where, when, and how such effectors evolved, given that they are often considered lineage-specific inventions<sup>79</sup>. Since pathogen effectors often lack conserved protein domains, it has been speculated that such effectors may have evolved de novo<sup>79</sup>. De novo

genes are defined to originate, at least in part, from non-coding DNA sequences, but the underlying evolutionary mechanisms remain poorly understood<sup>80,81</sup>. Our results suggest that a *Starship* contributed to the emergence of a virulence effector gene from a non-coding cargo sequence by the fusion with a section of a conserved effector gene. Given the strong association of *Starships* with plastic genomic regions enriched in (a)virulence effector genes, further elucidation of

**Fig. 3 | *Starships* show typical traits of the plastic genomic regions. a** *Verticillium* phylogeny as described in Fig. 1a. **b** Presence or absence of (a) virulence-associated genes in the 56 *Verticillium* strains. Square colors indicate *Starships* in which the genes were detected, while the other orthologs found in syntenic *Starship* regions are not colored. Circles indicate the presence of orthologs with 100% gene coverage and the fill color representing nucleotide identity. **c** Violin plots depicting the *in planta* expression induction of genes residing in the three genomic compartments in *V. dahliae* strain JR2. Induction levels are represented by fold-changes (FC) of gene expression in *planta* (*Arabidopsis thaliana*) versus *in vitro* (potato dextrose broth, PDB). Points indicate FC values for each gene ( $n = 450$  in *Starship* regions,  $n = 547$  in other adaptive genomic regions (AGRs), and  $n = 10,638$  in core regions), while blue crossbars indicate median values. Different letter labels indicate significant differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). **d** Violin plots depicting histone H3K27me3 levels for the three JR2 genomic compartments expressed as ChIP-Seq read counts per million (CPM) normalized by bin length. Each point indicates the value for a bin ( $\sim 10$  kb) ( $n = 192$  in *Starship* regions,  $n = 160$  in other AGRs, and  $n = 3145$  in core regions), while blue crossbars indicate median values. Different letter labels indicate significant differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). **e** Stacked bar plots depicting the transposable element (TE) density in the JR2 genomic compartments over 10 kb windows ( $n = 188$  in *Starship* regions,  $n = 153$  in other AGRs, and  $n = 3125$  in core regions). Different letter labels

indicate significant differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). **f** Proportion of expressed (transcripts per million (TPM)  $> 0$ ) and non-expressed (TPM = 0) TEs in the JR2 genomic compartments ( $n = 284$  in *Starship* regions,  $n = 202$  in other AGRs, and  $n = 1261$  in core regions). Different letter labels indicate significant differences (two-sided Fischer's test with Bonferroni correction, adjusted  $p < 0.05$ ). **g** Violin plots depicting the TE expression levels in the three JR2 genomic compartments for *V. dahliae* cultivated in PDB. Points indicate TPM values for individual TEs excluding non-expressed ones ( $n = 217$  in *Starship* regions,  $n = 78$  in other AGRs, and  $n = 650$  in core regions), while blue crossbars indicate median values. Different letter labels indicate significant differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). **h** Violin plots depicting repeat-induced point mutation (RIP) signature of TEs in the JR2 genomic compartments. Points indicate composite RIP index (CRI)<sup>48</sup> for individual TEs ( $n = 284$  in *Starship* regions,  $n = 202$  in other AGRs, and  $n = 1261$  in core regions), while blue crossbars indicate median values. Different letter labels indicate significant differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). **i** Locations of *Starships* and AGRs in the JR2 genome. See Fig. 2a legend for the details of symbols. Ribbons connect pairs of segmentally duplicated regions that share  $> 80\%$  nucleotide sequence identity over 10 kb, with a color representing identity. **j** Long-range chromatin interactions in the JR2 genome. Ribbons connect genomic regions that are separated in the genome but physically interact in nuclei by colors that represent the intensity.

*Starship*-mediated gene evolution can be fundamental to understanding how novel virulence genes evolve.

## Methods

### Genome assemblies and annotations

Genome sequences were collected from public databases or generated in this study (Supplementary Data 1, 9, 10). Fungal genomic DNA was extracted and sequenced using Oxford Nanopore Technologies as previously described<sup>34</sup>. Sequencing reads were assembled with Canu version 2.2<sup>82</sup>, and genome assembly quality was evaluated with BUSCO version 5.7.0 with the datasets eukaryota\_odb10 and glomerellales\_odb10<sup>83,84</sup>. The genome-wide average nucleotide identity (ANI) was calculated with FastANI version 1.33<sup>85</sup> that maps query genomic sequences fragmented to 3 kb to a reference genome and calculates the average of the maximum identity of the mapped regions.

Genes and TEs in *V. dahliae* strain JR2 were identified previously<sup>38,44</sup>. In all other genomes, repetitive sequences were predicted with RepeatModeler version 2.0.5<sup>86</sup> and were soft-masked with RepeatMasker version 4.1.5<sup>87</sup>. Genes were then predicted with the BRAKER version 3.0.8 pipeline C<sup>88</sup>, which trains GeneMark-EP+ version 4.72<sup>89</sup> and AUGUSTUS version 3.0.8<sup>90</sup> with information on splice sites, start, stop, and coding features using reference fungal proteins in OrthoDB version 11<sup>91</sup>. Orthologous gene groups were identified by eggNOG-mapper version 2.1.12<sup>92</sup> using eggNOG database version 5.0.2<sup>93</sup>. TEs were further characterized with the TE annotation pipeline EDTA version 2.2.1<sup>94</sup> with curated TEs of *V. dahliae* VdLs17<sup>62</sup>.

### Identification of *Starships* and *Starship* regions

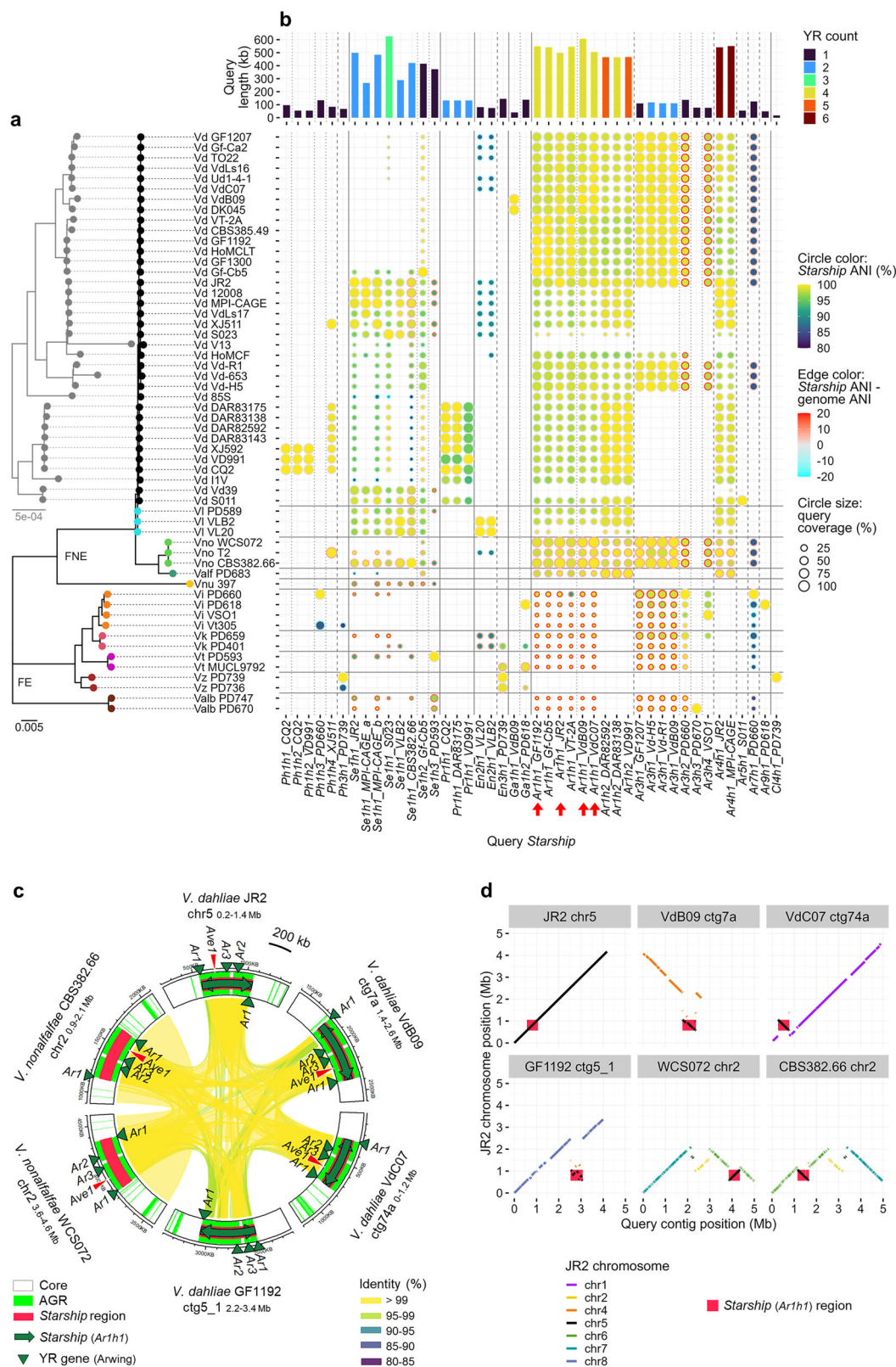
*Starships* and captain/captain-like tyrosine recombinase (YR) genes were identified with Starfish version 1.0.0 according to the standard workflow<sup>6</sup>. First, the captain candidate YR genes were identified de novo from the genomes using the Starfish gene finder module with MetaEuk version 6.a5d39d9<sup>95</sup> and HMMER version 3.3.2<sup>96</sup> based on sequence similarity to the Pezizomycotina captain/captain-like YR database attached to Starfish. The identified YR genes were grouped into families as defined previously<sup>6</sup> based on a similarity search with HMMER version 3.3.2<sup>96</sup>. The YR genes were then grouped into naves by clustering of *Verticillium* YRs using MMseqs2 version 14.7e284 easy-cluster<sup>97</sup> with thresholds of 50% amino acid sequence identity and 25% coverage. Each *Verticillium* YR nave was named with two letters of the YR family followed by an identifier number. Then, *Starship* candidates were identified via multiple sequence alignments among the 56

*Verticillium* genomes or among the 283 *Fusarium* genomes that were selected based on a relatively low number of contig/scaffold numbers, suggesting a relatively highly contiguous genome assembly (Supplementary Data 2, 10), by the Starfish element finder module using BLAST version 2.12.0+<sup>98</sup> and MUMmer version 4.0.0rc1<sup>99</sup>. After the initial screening by Starfish, confident *Starships* were curated by (1) removing *Starship* candidates that lack flanking synteny over 20 kb against orthologous sites without *Starships* through manual inspection with Starfish pairViz, (2) removing *Starship* candidates that contain serial N stretches ( $N \geq 3$ ) in scaffolds, and (3) unifying redundantly detected *Starships* due to the presence of YR genes on both strands around the 5' ends. The *Starships* were grouped into haplotypes based on the captain navis and nucleotide  $k$ -mer similarities detected by sourmash version 4.8.3<sup>100</sup> with Starfish sim with default  $k$ -mer size 510<sup>6,9</sup>, followed by the clustering using mcl version 14-137<sup>101</sup> with minimal similarity threshold 0.05. Each *Verticillium* *Starship* haplotype was named after the captain navis ID followed by a suffix. Genomic regions orthologous to *Starship* insertion sites across *Verticillium* strains (Supplementary Fig. S7) were identified by the Starfish region finder module based on the presence of low copy number (up to 5) eggNOG orthologs.

*Starship* regions were annotated by two approaches. First, regions downstream of YRs were annotated as *Starship* regions using Starfish extend<sup>6</sup> with default settings using BLAST version 2.12.0+<sup>98</sup> based on similarity against curated *Starships* (Supplementary Data 12). Second, other *Starship* regions were identified irrespective of the presence of YR genes based on synteny against curated *Starships* through the alignments with nucmer of MUMmer version 4.0.0rc1 with options maxmatch and minimal alignment length 10,000, followed by filtering using delta-filter with the threshold of  $> 80\%$  nucleotide identity<sup>99</sup> (Supplementary Data 13). *Starships* identified in this study (Supplementary Data 3, 11) were used as reference. The *Starship* regions identified by the two approaches were merged with bedtools version 2.30.0<sup>102</sup> to determine the final coordinates of *Starship* regions (Supplementary Data 5). The pairwise sequence alignments between each *Starship* and syntenic *Starship* regions in each genome (Fig. 4b) were also performed using nucmer with the same settings.

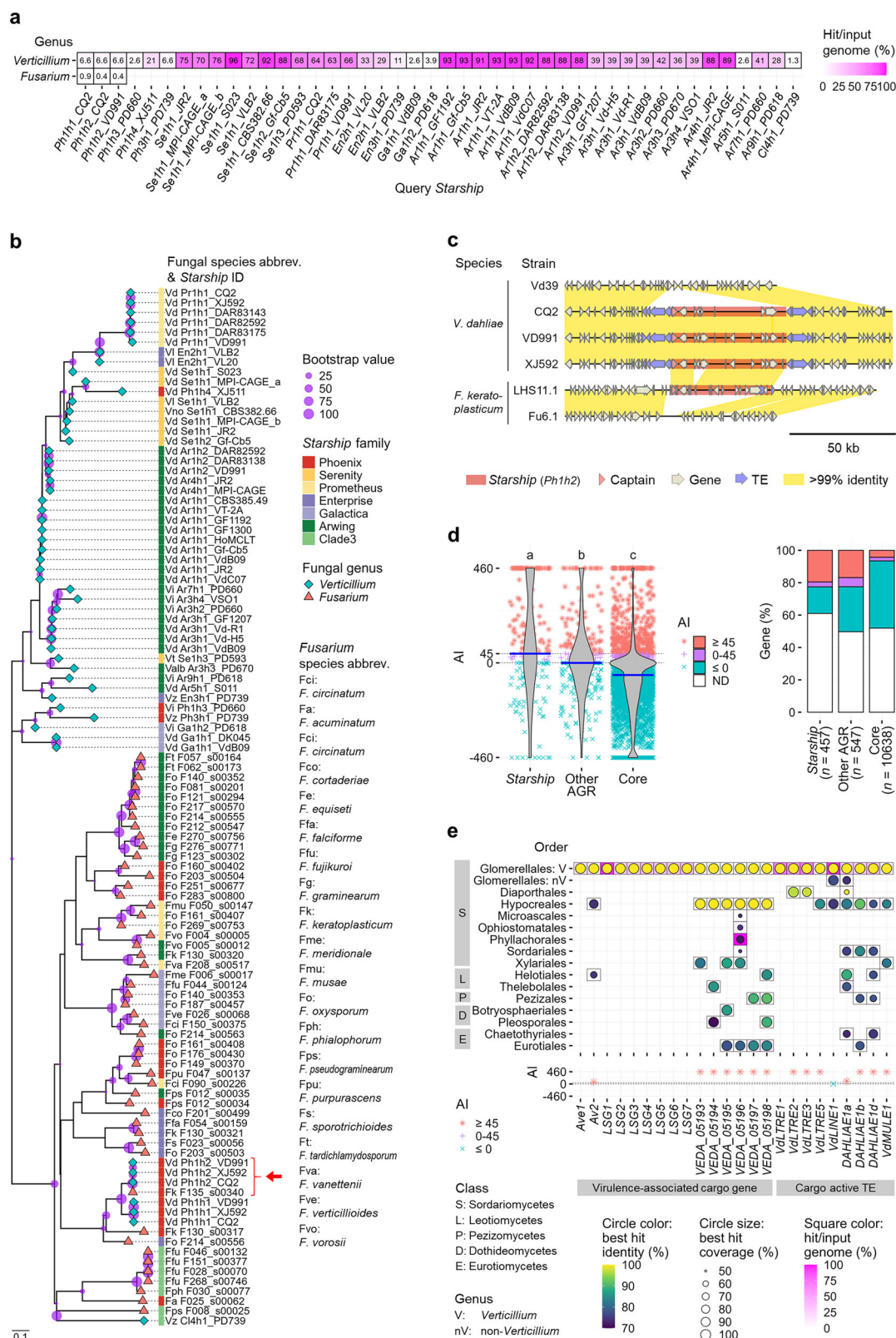
The genomic compartments of *V. dahliae* strains JR2 and VdLs17 were assigned based on the overlap of *Starships* and *Starship* regions with the previously assigned AGRs, core genomic regions<sup>30</sup>, and centromeres<sup>103</sup>. AGRs in the JR2 genome were identified by chromatin profiling<sup>30</sup>, while AGRs in the other strains refer to regions syntenic to AGRs in JR2 and regions absent in JR2<sup>42</sup>. To identify AGRs in strains





**Fig. 4 | Starship mobility among *Verticillium* genomes.** **a** *Verticillium* phylogeny as described in Fig. 1a. **b** Occurrence of diverse *Starships* across the *Verticillium* genus. Columns indicate different *Starships* with their length and number of captain and captain-like tyrosine recombinase (YR) genes. The heatmap indicates the query coverage, Average Nucleotide Identity (ANI) of *Starship* regions (*Starship* ANI), and the difference between *Starship* ANI and genome-wide ANI for each pairwise strain comparison to detect potential horizontal transfers. Red arrows indicate the *Starship* highlighted in (c). **c** Circular plot depicting the synteny among

regions of four *V. dahliae* and two *V. nonalfalfae* strains harboring orthologous *Ar1h1* *Starships*. See Fig. 2a legend for the details of symbols. **d** Synteny plots between the *V. dahliae* JR2 genome and the *Ar1h1* *Starship* regions in (c). Each plot indicates the synteny between the JR2 chromosomes (Y-axis) and each chromosome (chr) or contig (ctg) containing the *Ar1h1* *Starship* (X-axis), with diagonal lines colored by syntenic JR2 chromosome. Red background indicates the coordinates of *Ar1h1* *Starship*.



other than JR2, genome alignments were performed with nucmer of MUMmer version 4.0.0rc1<sup>99</sup>, followed by filtering of alignments with >80% nucleotide identity over 1 kb. The regions aligned to the JR2 AGRs and not aligned to the JR2 genome were merged with bedtools version 2.30.0<sup>102</sup> and then filtered by the length threshold of >1 kb to determine the final AGRs. The “Starship” compartment refers to *Starships* and additional *Starship* regions, “other AGR” refers to AGRs that

do not belong to *Starship* regions, “centromere” refers to previously identified centromeres, and “core” refers to core genomic regions that do not belong to *Starship* regions or centromeres. Genes and TEs in the JR2 genome were assigned to genomic compartments based on the location of their midpoints. TE counts over 10 kb windows in each genomic compartment were calculated using bedtools version 2.30.0 intersect<sup>102</sup>. Compartments smaller than 10 kb were omitted from the



**Fig. 5 | Starship dynamics among fungal orders.** **a** Occurrence of *Verticillium* *Starships* across 10,113 Pezizomycotina genomes. Cell colors and labels denote the percentage of genomes with hits (e-value < 0.05 and total query coverage >50% or >30 kb) in each genus. **b** Phylogeny of *Starships* that occur in the *Verticillium* and *Fusarium* genera based on *k*-mer similarity. Scale bar indicates the Mash distance that represents the *k*-mer difference<sup>140</sup>. Circles at the nodes represent bootstrap values for 1000 iterations. Red arrows indicate the *Starship* highlighted in (c). **c** Similarity of *PhlH2* *Starships* and surrounding regions between *V. dahliae* and *F. keratoplasticum*. **d** Possibility of horizontal gene transfer (HGT) between *Verticillium* and fungi belonging to the other orders of Pezizomycotina for genes residing in the three genomic compartments in *V. dahliae* strain JR2. High Alien index (AI) values ( $\geq 45$ ) suggest HGT, while moderate values ( $>0$  and  $<45$ ) suggest a

weak HGT possibility. Points in the left violin plots indicate AI values of individual genes ( $n = 457$  (including captain and captain-like genes predicted de novo) in *Starship* regions,  $n = 547$  in other adaptive genomic regions (AGRs), and  $n = 10638$  in core regions), while blue crossbars indicate median values. Different letter labels indicate significant differences (two-sided Dunn's test, adjusted  $p < 0.05$ ). The right bar plots depict the percentage of genes with high, moderate, and low AI values. ND indicates that AI was not determined because of lacking hits in both ingroup and outgroup. **e** Occurrence of homologs of *Verticillium* *Starship* cargo genes and transposable elements (TEs) across 10,113 Pezizomycotina genomes. Square colors represent the percentage of genomes with hits (e-value < 0.05, over 70% nucleotide identity over 50% coverage) in each group. Circles indicate the coverage and identity of the best hits. The lower panel indicates AI values as described in (d).

TE density analysis. The composite RIP index (CRI) of TEs was calculated as previously described<sup>48</sup>. Briefly, CRI was determined by subtracting the RIP substrate index ((CpA + TpG)/(ApC + GpT)) from the RIP product index (TpA/ApT), as CpA dinucleotides are preferentially targeted by RIP.

### Analysis of genomic synteny and variations

Genomic synteny and variations were identified with MUMmer version 4.0.0rc1<sup>99</sup>. Specifically, segmental duplications in the JR2 genome were identified by genome self-alignment using nucmer with options maxmatch, nosimplify, and minimal alignment length 10,000, followed by filtering with delta-filter. Inter-genomic synteny and variations were identified by pairwise genome alignments with nucmer with option minimal alignment length 10,000, followed by filtering and analysis by dnadiff. The genome-wide alignment coverage was calculated by integrating all fragmented hits. Among the structural variations (SVs) shown in this study, insertion/deletion (INDEL) corresponds to “GAP” and “BRK” in MUMmer4, while translocation (TRA) corresponds to “JMP” and “SEQ”. SV frequency was calculated by dividing the SV counts by the aligned length for each genomic compartment or genome. SV fold-enrichment was calculated by dividing the frequency in each genomic compartment by the genome-wide frequency.

### Transcriptome, epigenome, and 3D genome analyses

Global gene expression in *V. dahliae* strain JR2 was analyzed using the previous RNA sequencing (RNA-Seq) read data<sup>104</sup> obtained from the NCBI short read archive (SRA) under the accession numbers listed in Supplementary Data 14. RNA-Seq reads were filtered with fastp version 0.19.5<sup>105</sup> and mapped to the unmasked JR2 genome with STAR version 2.7.10a<sup>106</sup> with previously applied options to allow multiple mapped reads for TE expression measurement<sup>38</sup>. The mapped reads were counted for each gene and TE by Tetranscripts version 2.2.1 with mode multi for 1,000 iterations to allow fractional counting of multi-mapped reads<sup>107</sup>. Fold change (FC) of gene expression was analyzed by DESeq2 version 1.42.1<sup>108</sup>. Genes and TEs with no sequencing reads in all samples were excluded from the DESeq2 analysis. Transcripts per million (TPM) were calculated using the established formula<sup>109</sup>. Mean TPM values of three biological replicates for individual genes/TEs were used for the data visualization and statistics.

Global histone H3K27me3 modifications in *V. dahliae* strain JR2 were analyzed using the previous chromatin immunoprecipitation sequencing (ChIP-Seq) data<sup>30</sup> obtained from NCBI SRA under the accession numbers listed in Supplementary Data 14. ChIP-Seq reads were filtered with fastp version 0.19.5<sup>105</sup> and mapped to the JR2 genome masked as previously described<sup>30</sup> using BWA version 0.7.17 with the BWA-MEM algorithm<sup>110</sup>. The mapped reads were sorted and indexed with samtools version 1.10<sup>111</sup> and counted with featureCounts version 2.0.1<sup>112</sup> with the option to count multi-mapped reads fractionally for each genomic compartment over 10 kb windows. The normalized counts per million (CPM) were calculated using the TPM formula<sup>109</sup> by replacing transcripts with bins. The mean normalized

CPM values of two biological replicates for respective bins were used for the data visualization and statistics.

Long-range chromatin interactions in *V. dahliae* strain JR2 were previously identified through chromatin conformation capture (Hi-C)<sup>42</sup>. The previously analyzed data were directly used for visualization with the *Starship* positions.

### Local similarity searches

Local similarity searches were performed with BLAST version 2.15.0+ or 2.16.0+<sup>98</sup>. The nucleotide-to-nucleotide searches were performed with the blastn algorithm with options word size 11 and e-value 0.05. The protein-to-translated nucleotide searches were performed with the tblastn algorithm with an e-value threshold of 0.05. The query *Verticillium* gene and TE sequences were obtained from GenBank and FungiDB under the accession numbers listed in Supplementary Data 15, and genes including introns were used as queries. The sequences and metadata of other virulence-associated proteins were obtained from PHI-base version 4.17<sup>49</sup>. Pezizomycotina genomes were obtained from whole-genome shotgun (WGS) sequences in GenBank under the accession numbers listed in Supplementary Data 9. Fungal taxonomic data were obtained from the National Center for Biotechnology Information (NCBI) Taxonomy database. The coordinates of *Au2* orthologs (Supplementary Data 16) were identified by a blastn search of the Pezizomycotina genomes with *V. dahliae* *Au2* or the *Au2* ortholog of *Fusarium phyllophilum* (*FpAu2*) under the accession numbers in Supplementary Data 15, followed by the selection of hits with >70% identity and >80% coverage.

Alien index (AI) values were calculated by the formula  $AI = \log((\text{best e-value for ingroup}) + e-200) - \log((\text{best e-value for outgroup}) + e-200)$ <sup>61</sup> using hits with >50% coverage against *Verticillium* gene/TE queries by the blastn search. The e-value for no hits was set to 1 for queries with hits in either ingroup or outgroup, whereas AI values for queries with no hits in both ingroup and outgroup were not determined. The 332 non-*Verticillium* genomes of the order Glomerellales were used as ingroup, while the 9641 genomes of the remaining 74 orders of Pezizomycotina were used as outgroup (Supplementary Data 9).

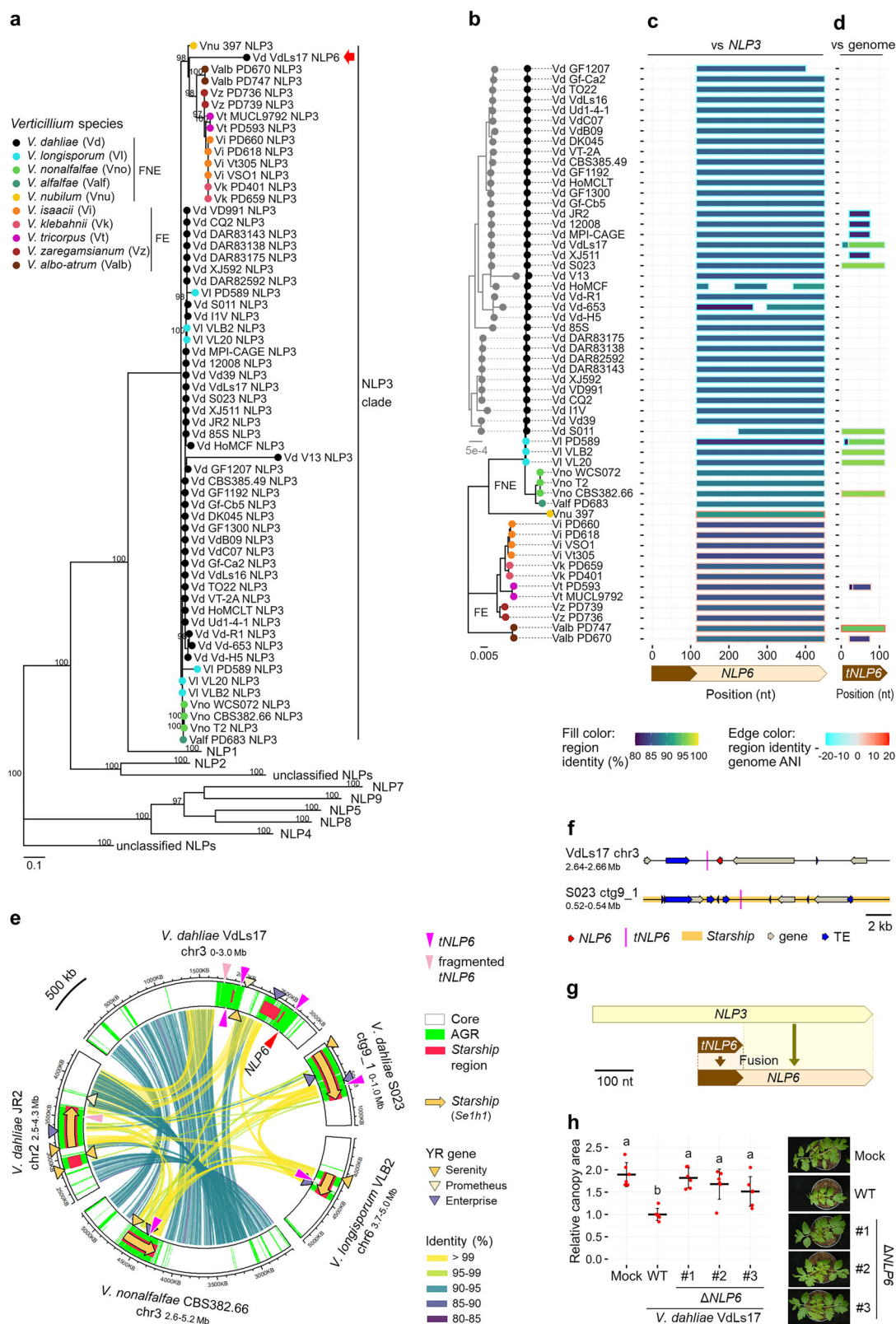
### Phylogenetic analyses

*Verticillium* phylogeny was inferred with REALPHY version 1.13<sup>113</sup>. Briefly, genome sequences were fragmented into overlapping 50 bp subsequences and mapped to the JR2 genome with Bowtie version 2.2.5<sup>114</sup>. Based on single-nucleotide polymorphisms in the aligned regions, the maximum likelihood phylogenetic tree was built with PhyML version 3.3.20220408<sup>115</sup>.

*Starship* phylogeny was inferred by *k*-mer comparisons with mashree version 1.4.6 with default *k*-mer size of 21, accuracy option mindepth 0, and bootstrapping for 1,000 iterations<sup>116</sup>.

For the phylogenetic analyses of *Au2*, nucleotide sequences of *Au2* orthologs were extracted from the Pezizomycotina genomes at the coordinates listed in Supplementary Data 16 using SeqKit version 2.3.0 subseq<sup>117</sup>. The nucleotide sequences of *Au2* were aligned by





MAFFT versions 7.511 or 7.526 with the L-INS-i method that iteratively refines local alignments<sup>118</sup>, followed by removal of ambiguously aligned sites with trimAl version 1.4.rev15 with option strict<sup>119</sup>. After multiple sequence alignments, maximum likelihood phylogeny was inferred by IQ-TREE version 2.0.3<sup>120</sup> with the best substitution model TIM2e + G4 suggested by ModelFinder<sup>121</sup> and the ultrafast bootstrap approximation for 1000 iterations<sup>122</sup>.

For the phylogenetic analysis of NLPs, putative *Verticillium* proteins that contain an NPP1 domain<sup>64</sup> were identified based on the homology to reference sequences in the Pfam database<sup>123</sup> (Pfam accession PF05630.16) using HMMER version 3.3.2<sup>96</sup>. Putative secretion signals were identified with SignalP 6.0<sup>124</sup> or SignalP 3.0<sup>125</sup>. The deduced NLP amino acid sequences were aligned and trimmed as described for *Av2*. The phylogeny of NLPs was also inferred as

**Fig. 6 | Starships contributed to de novo formation of the virulence gene *NLP6*.** **a** Phylogeny of necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs) in 56 strains across the *Verticillium* genus with all NLP clades collapsed except for the NLP3 clade. The red arrow points to NLP6. Scale bar indicates amino acid substitutions per site. Bootstrap values (>95%) for 1000 iterations are shown at the nodes. Circle colors indicate *Verticillium* spp. from which each NLP homolog was derived. Flavexudans (FE) and Flavonoxudans (FNE) indicate the two clades within the *Verticillium* genus<sup>28</sup>. **b** *Verticillium* phylogeny as described in Fig. 1a. Circle colors indicate *Verticillium* spp. with the same color coding as in (a). **c, d** Coverage plots for alignments of *NLP6* with *NLP3* genes (c) and for alignments of a truncated *NLP6* (tNLP6; 1–114 nt of *NLP6*) with genome sequences (d). The X-axes represent the

nucleotide positions in *NLP6* or tNLP6 of *V. dahliae* VdLs17. Bar colors indicate the identity of the region and its difference from genome-wide ANI between each *Verticillium* strain and VdLs17. **e** Location of tNLP6 in *Starship* in three *Verticillium* species. See Fig. 2a legends for the details of symbols. **f** Location of tNLP6 in non-coding regions of *V. dahliae* strains. **g** Proposed model for *NLP6* evolution. **h** Symptoms of tomato plants inoculated with wild-type (WT) and three independent *NLP6* deletion ( $\Delta$ ) lines of *V. dahliae* strain VdLs17 at 14 days post inoculation. Points indicate relative values for individual plants ( $n = 6$ ), divided by the mean of WT-inoculated plants. Crossbars and error bars indicate mean  $\pm$  standard deviation. Different letter labels indicate significant differences (two-sided Tukey's test, adjusted  $p < 0.05$ ).

described for *Av2* with the different best substitution model VT + F + I + G4. NLP orthologs were numbered by the monophyletic relationship with NLP1 to NLP9 of *V. dahliae* strain VdLs17<sup>39,63</sup> in the GenBank RefSeq database under the accession numbers listed in Supplementary Data 15.

### Targeted deletion of NLPs from the *Verticillium dahliae* genome

To generate *NLP6* and *NLP3* deletion constructs, flanking sequences of its coding sequence were amplified from genomic DNA of *V. dahliae* strains VdLs17 and JR2, respectively, using primers listed in Supplementary Data 17. The amplified products were cloned into pRF-HU2 as described previously<sup>126</sup>, and subsequent *Agrobacterium tumefaciens*-mediated transformation of *V. dahliae* was performed as described previously<sup>127</sup>. Transformants were selected on PDA supplemented with cefotaxime (Duchefa, Haarlem, The Netherlands) at 200  $\mu$ g/ml and hygromycin (Duchefa) at 50  $\mu$ g/ml, and homologous recombination was PCR-verified. For genetic complementation, the coding sequence of *NLP6* was cloned into the pFBT005 vector as previously described<sup>128</sup>, after which the *NLP6* deletion mutants were transformed using the *A. tumefaciens*-mediated transformation method described above.

Pathogenicity assays were performed on ten-day-old tomato seedlings (MoneyMaker) plants using root-dip inoculation as previously described<sup>129</sup>. Disease symptoms were scored up to 14 dpi, pictures were taken, and ImageJ was used to determine canopy areas while fungal colonization was determined with real-time PCR. To this end, stem sections were taken at the height of the first internode, flash-frozen in liquid nitrogen, ground to powder, and genomic DNA was isolated. Real-time PCR was performed with a quantitative PCR core kit for SYBR Green I (Eurogentec, Seraing, Belgium) on an ABI7300 PCR machine (Applied Biosystems, Foster City, CA, U.S.A.). The *V. dahliae* internal transcribed spacer (ITS) levels were used relative to tomato ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) levels to quantify fungal colonization of tomato plants<sup>36</sup>.

### Statistics

Data were analyzed with R version 4.4.2<sup>130</sup> and the R package “tidyverse” version 2.0.0<sup>131</sup>. The normality and homoscedasticity of data were tested by the Shapiro–Wilk test and Bartlett's test, respectively, with the R base package “stats” version 4.3.1<sup>130</sup>. The multiple comparisons of data with non-normal distributions and unequal variances ( $p < 0.05$ ) were performed by Dunn's test with the Bonferroni correction with the R package “dunn.test” version 1.3.6<sup>132</sup>. The multiple comparisons of data with normal distributions and equal variances ( $p < 0.05$ ) were performed by Tukey's test with the R package “multcomp” version 1.4.26<sup>133</sup>. The multiple comparisons of the proportions of two categorical variables were performed by Fisher's test with the Bonferroni correction with the R package “RVAideMemoire” version 0.9-83-11<sup>134</sup>. The exact adjusted  $p$ -values of multiple comparisons are provided in the Source Data files. The Spearman's rank correlation coefficient was calculated with the R base package “stats” version 4.3.1<sup>130</sup>.

### Data visualization

Data were visualized with the R packages. Circular and linear plots of genomic regions were generated with “circlize” version 0.4.16<sup>135</sup> and “gggenomes” version 1.0.0<sup>136</sup> or “genoPlotR” version 0.8.11<sup>137</sup>, respectively. Phylogenetic trees were visualized with “ggtree” version 3.10.1<sup>138</sup>. Other plots were generated with “ggplot2” version 3.5.1<sup>139</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The genomes assembled in this study and previous studies<sup>57,103</sup> have been submitted to NCBI under the BioProject accession PRJNA1253319. Genome annotation files and genome assemblies with complex gap information<sup>103</sup> have been deposited at Zenodo (<https://doi.org/10.5281/zenodo.15450312>). Other genome sequence, RNA-Seq, and ChIP-Seq data used in this study are available in the NCBI database under the accession numbers listed in Supplementary Data 1, 9, 10, 14–16. Source data are provided with this paper.

### References

1. Fedoroff, N. V. Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–767 (2012).
2. Wicker, T. et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
3. Arkhipova, I. R. & Yushenova, I. A. Giant transposons in eukaryotes: is bigger better?. *Genome Biol. Evol.* **11**, 906–918 (2019).
4. Inoue, Y. & Takeda, H. *Teratorn* and its relatives – a cross-point of distinct mobile elements, transposons and viruses. *Front. Vet. Sci.* **10**, 1158023 (2023).
5. Urquhart, A., Vogan, A. A. & Gluck-Thaler, E. *Starships*: a new frontier for fungal biology. *Trends Genet.* **40**, 1060–1073 (2024).
6. Gluck-Thaler, E. & Vogan, A. A. Systematic identification of cargo-mobilizing genetic elements reveals new dimensions of eukaryotic diversity. *Nucleic Acids Res.* **52**, 5496–5513 (2024).
7. Gluck-Thaler, E. et al. Giant *Starship* elements mobilize accessory genes in fungal genomes. *Mol. Biol. Evol.* **39**, msac109 (2022).
8. Urquhart, A. S., Vogan, A. A., Gardiner, D. M. & Idnurm, A. *Starships* are active eukaryotic transposable elements mobilized by a new family of tyrosine recombinases. *Proc. Natl Acad. Sci. USA* **120**, e2214521120 (2023).
9. Gluck-Thaler, E. et al. Giant transposons promote strain heterogeneity in a major fungal pathogen. *mBio* **16**, e01092-25 (2025).
10. Bucknell, A. H. & McDonald, M. C. That's no moon, it's a *Starship*: Giant transposons driving fungal horizontal gene transfer. *Mol. Microbiol.* **120**, 555–563 (2023).
11. Urquhart, A. S., Chong, N. F., Yang, Y. & Idnurm, A. A large transposable element mediates metal resistance in the fungus *Paecilomyces variotii*. *Curr. Biol.* **32**, 937–950.e5 (2022).

12. Gourlie, R. et al. The pangenome of the wheat pathogen *Pyrrenophora tritici-repentis* reveals novel transposons associated with necrotrophic effectors *ToxA* and *ToxB*. *BMC Biol.* **20**, 239 (2022).
13. Urquhart, A. S., Gluck-Thaler, E. & Vogan, A. A. Gene acquisition by giant transposons primes eukaryotes for rapid evolution via horizontal gene transfer. *Sci. Adv.* **10**, eadp8738 (2024).
14. Peck, L. D. et al. Horizontal transfers between fungal *Fusarium* species contributed to successive outbreaks of coffee wilt disease. *PLoS Biol.* **22**, e3002480 (2024).
15. Bucknell, A. et al. *Sanctuary*: a *Starship* transposon facilitating the movement of the virulence factor *ToxA* in fungal wheat pathogens. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.03.04.583430> (2024).
16. O'Donnell, S., Rezende, G., Vernadet, J.-P., Snirc, A. & Ropars, J. Harboring *Starships*: the accumulation of large horizontal gene transfers in domesticated and pathogenic fungi. *Genome Biol. Evol.* <https://doi.org/10.1093/gbe/evaf125> (2025).
17. Dangl, J. L. & Jones, J. D. G. A common immune response node in diverse plants. *Science* **386**, 1344–1346 (2024).
18. Cook, D. E., Mesarich, C. H. & Thomma, B. P. H. J. Understanding plant immunity as a surveillance system to detect invasion. *Annu. Rev. Phytopathol.* **53**, 541–563 (2015).
19. Möller, M. & Stukenbrock, E. H. Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* **15**, 756–771 (2017).
20. Raffaele, S. & Kamoun, S. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* **10**, 417–430 (2012).
21. Dong, S., Raffaele, S. & Kamoun, S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.* **35**, 57–65 (2015).
22. Seidl, M. F., Cook, D. E. & Thomma, B. P. H. J. Chromatin biology impacts adaptive evolution of filamentous plant pathogens. *PLoS Pathog.* **12**, e1005920 (2016).
23. Seidl, M. F. & Thomma, B. P. H. J. Transposable elements direct the coevolution between plants and microbes. *Trends Genet.* **33**, 842–851 (2017).
24. Kramer, H. M., Cook, D. E., Seidl, M. F. & Thomma, B. P. H. J. Epigenetic regulation of nuclear processes in fungal plant pathogens. *PLoS Pathog.* **19**, e1011525 (2023).
25. Frantzeskakis, L., Kusch, S. & Panstruga, R. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol. Plant Pathol.* **20**, 3–7 (2019).
26. Torres, D. E., Oggenfuss, U., Croll, D. & Seidl, M. F. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol. Rev.* **34**, 136–143 (2020).
27. Fradin, E. F. & Thomma, B. P. H. J. Physiology and molecular aspects of *Verticillium* wilt diseases caused by *V. dahliae* and *V. albo-atrum*. *Mol. Plant Pathol.* **7**, 71–86 (2006).
28. Inderbitzin, P. et al. Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. *PLoS ONE* **6**, e28341 (2011).
29. Klimes, A., Dobinson, K. F., Thomma, B. P. H. J. & Klosterman, S. J. Genomics spurs rapid advances in our understanding of the biology of vascular wilt pathogens in the genus *Verticillium*. *Annu. Rev. Phytopathol.* **53**, 181–198 (2015).
30. Cook, D. E., Kramer, H. M., Torres, D. E., Seidl, M. F. & Thomma, B. P. H. J. A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen. *eLife* **9**, e62208 (2020).
31. de Jonge, R. et al. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc. Natl Acad. Sci. USA* **109**, 5110–5115 (2012).
32. de Jonge, R. et al. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res.* **23**, 1271–1282 (2013).
33. Snelders, N. C. et al. Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nat. Plants* **6**, 1365–1374 (2020).
34. Chavarro-Carrero, E. A. et al. Comparative genomics reveals the *in planta*-secreted *Verticillium dahliae* Av2 effector protein recognized in tomato plants that carry the V2 resistance locus. *Environ. Microbiol.* **23**, 1941–1958 (2021).
35. Kombrink, A. et al. *Verticillium dahliae* LysM effectors differentially contribute to virulence on plant hosts. *Mol. Plant Pathol.* **18**, 596–608 (2017).
36. Snelders, N. C. et al. A highly polymorphic effector protein promotes fungal virulence through suppression of plant-associated Actinobacteria. *N. Phytologist* **237**, 944–958 (2023).
37. Faino, L. et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* **26**, 1091–1100 (2016).
38. Torres, D. E., Thomma, B. P. H. J. & Seidl, M. F. Transposable elements contribute to genome dynamics and gene expression variation in the fungal plant pathogen *Verticillium dahliae*. *Genome Biol. Evol.* **13**, evab135 (2021).
39. Klosterman, S. J. et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog.* **7**, e1002137 (2011).
40. Kramer, H. M., Cook, D. E., van den Berg, G. C. M., Seidl, M. F. & Thomma, B. P. H. J. Three putative DNA methyltransferases of *Verticillium dahliae* differentially contribute to DNA methylation that is dispensable for growth, development and virulence. *Epigenetics Chromatin* **14**, 21 (2021).
41. Kramer, H. M., Seidl, M. F., Thomma, B. P. H. J. & Cook, D. E. Local rather than global H3K27me3 dynamics are associated with differential gene expression in *Verticillium dahliae*. *mBio* **13**, e03566–21 (2022).
42. Torres, D. E. et al. Implications of the three-dimensional chromatin organization for genome evolution in a fungal plant pathogen. *Nat. Commun.* **15**, 1701 (2024).
43. Shi-Kunne, X., Faino, L., van den Berg, G. C. M., Thomma, B. P. H. J. & Seidl, M. F. Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene loss. *Environ. Microbiol.* **20**, 1362–1373 (2018).
44. Faino, L. et al. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *mBio* <https://doi.org/10.1128/mbio.00936-15> (2015).
45. Kraege, A. et al. Undermining the cry for help: the phytopathogenic fungus *Verticillium dahliae* secretes an antimicrobial effector protein to undermine host recruitment of antagonistic *Pseudomonas* bacteria. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.06.09.658588> (2025).
46. Freitag, M., Williams, R. L., Kothe, G. O. & Selker, E. U. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc. Natl Acad. Sci. USA* **99**, 8802–8807 (2002).
47. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20**, 417–423 (2004).
48. Lewis, Z. A. et al. Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. *Genome Res.* **19**, 427–437 (2009).
49. Urban, M. et al. PHI-base in 2022: a multi-species phenotype database for pathogen–host interactions. *Nucleic Acids Res.* **50**, D837–D847 (2022).
50. Chen, J.-Y. et al. Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in *Verticillium dahliae*



- and evidence of horizontal gene transfer from *Fusarium*. *N. Phytologist* **217**, 756–770 (2018).
51. Harting, R. et al. A 20-kb lineage-specific genomic region tames virulence in pathogenic amphidiploid *Verticillium longisporum*. *Mol. Plant Pathol.* **22**, 939–953 (2021).
  52. Caracul, Z. et al. The pH signalling transcription factor PacC controls virulence in the plant pathogen *Fusarium oxysporum*. *Mol. Microbiol.* **48**, 765–779 (2003).
  53. Ortoneda, M. et al. *Fusarium oxysporum* as a multihost model for the genetic dissection of fungal virulence in plants and mammals. *Infect. Immun.* **72**, 1760–1766 (2004).
  54. Nguyen, Q. B. et al. Systematic functional analysis of calcium-signalling proteins in the genome of the rice-blast fungus, *Magnaporthe oryzae*, using a high-throughput RNA-silencing system. *Mol. Microbiol.* **68**, 1348–1365 (2008).
  55. Zhang, J., Jin, K. & Xia, Y. Contributions of  $\beta$ -tubulin to cellular morphology, sporulation and virulence in the insect-fungal pathogen, *Metarhizium acridum*. *Fungal Genet. Biol.* **103**, 16–24 (2017).
  56. Kodama, S. et al. The morphogenesis-related NDR kinase pathway of *Colletotrichum orbiculare* is required for translating plant surface signals into infection-related morphogenesis and pathogenesis. *PLoS Pathog.* **13**, e1006189 (2017).
  57. Depotter, J. R. L. et al. Dynamic virulence-related regions of the plant pathogenic fungus *Verticillium dahliae* display enhanced sequence conservation. *Mol. Ecol.* **28**, 3482–3495 (2019).
  58. Ravenhall, M., Škunca, N., Lassalle, F. & Dessimoz, C. Inferring Horizontal Gene Transfer. *PLoS Comput. Biol.* **11**, e1004095 (2015).
  59. Vogan, A. A. et al. The *Enterprise*, a massive transposon carrying *Spok* meiotic drive genes. *Genome Res.* **31**, 789–798 (2021).
  60. Hill, R. et al. *Starship* giant transposable elements cluster by host taxonomy using *k*-mer-based phylogenetics. *G3* **15**, jkaf082 (2025).
  61. Gladyshev, E. A., Meselson, M. & Arkhipova, I. R. Massive horizontal gene transfer in *Bdelloid* rotifers. *Science* **320**, 1210–1213 (2008).
  62. Amyotte, S. G. et al. Transposable elements in phytopathogenic *Verticillium* spp.: insights into genome evolution and inter- and intra-specific diversification. *BMC Genomics* **13**, 314 (2012).
  63. Santhanam, P. et al. Evidence for functional diversification within a fungal NEP1-like protein family. *Mol. Plant Microbe Interact.* **26**, 278–286 (2013).
  64. Seidl, M. F. & Ackerveken, G. V. den. Activity and phylogenetics of the broadly occurring family of microbial Nep1-like proteins. *Annu. Rev. Phytopathol.* **57**, 367–386 (2019).
  65. Haridas, S. et al. T-toxin virulence genes: unconnected dots in a sea of repeats. *mBio* **14**, e00261-23 (2023).
  66. Hartmann, F. E. et al. An inversion polymorphism under balancing selection, involving giant mobile elements, in an invasive fungal pathogen. *Mol. Biol. Evol.* **42**, msaf026 (2025).
  67. Hill, R. et al. Evolutionary genomics reveals variation in structure and genetic content implicated in virulence and lifestyle in the genus *Gaeumannomyces*. *BMC Genomics* **26**, 239 (2025).
  68. Tralamazza, S. M., Gluck-Thaler, E., Feurtey, A. & Croll, D. Copy number variation introduced by a massive mobile element facilitates global thermal adaptation in a fungal wheat pathogen. *Nat. Commun.* **15**, 5728 (2024).
  69. Seidl, M. F. & Thomma, B. P. H. J. Sex or no sex: evolutionary adaptation occurs regardless. *Bioessays* **36**, 335–345 (2014).
  70. Fitzpatrick, D. A. Horizontal gene transfer in fungi. *FEMS Microbiol. Lett.* **329**, 1–8 (2012).
  71. Soanes, D. & Richards, T. A. Horizontal gene transfer in eukaryotic plant pathogens. *Annu. Rev. Phytopathol.* **52**, 583–614 (2014).
  72. Patron, N. J. et al. Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. *BMC Evol. Biol.* **7**, 174 (2007).
  73. Khaldi, N. & Wolfe, K. H. Evolutionary origins of the fumonisin secondary metabolite gene cluster in *Fusarium verticillioides* and *Aspergillus niger*. *Int. J. Evol. Biol.* **2011**, 423821 (2011).
  74. Slot, J. C. & Rokas, A. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Curr. Biol.* **21**, 134–139 (2011).
  75. Campbell, M. A., Rokas, A. & Slot, J. C. Horizontal transfer and death of a fungal secondary metabolic gene cluster. *Genome Biol. Evol.* **4**, 289–293 (2012).
  76. Qiu, H., Cai, G., Luo, J., Bhattacharya, D. & Zhang, N. Extensive horizontal gene transfers between plant pathogenic fungi. *BMC Biol.* **14**, 41 (2016).
  77. Hiruma, K. et al. A fungal sesquiterpene biosynthesis gene cluster critical for mutualist-pathogen transition in *Colletotrichum tofielliae*. *Nat. Commun.* **14**, 5288 (2023).
  78. Kobayashi, N. et al. Horizontally transferred DNA in the genome of the fungus *Pyricularia oryzae* is associated with repressive histone modifications. *Mol. Biol. Evol.* **40**, msad186 (2023).
  79. Fouché, S., Plissonneau, C. & Croll, D. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr. Opin. Microbiol.* **46**, 34–42 (2018).
  80. McLysaght, A. & Hurst, L. D. Open questions in the study of de novo genes: what, how and why. *Nat. Rev. Genet.* **17**, 567–578 (2016).
  81. Oss, S. B. V. & Carvunis, A.-R. De novo gene birth. *PLoS Genet.* **15**, e1008160 (2019).
  82. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
  83. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
  84. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
  85. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
  86. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
  87. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker open-4.0. <http://www.repeatmasker.org> (2013).
  88. Gabriel, L. et al. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).
  89. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom. Bioinform.* **2**, lqaa026 (2020).
  90. Keller, O., Kollmar, M., Stanke, M. & Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757–763 (2011).
  91. Kuznetsov, D. et al. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **51**, D445–D451 (2023).
  92. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation,

- orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
93. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
  94. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
  95. Levy, Karin, E., Mirdita, M. & Söding, J. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome* **8**, 48 (2020).
  96. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
  97. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
  98. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
  99. Marçais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
  100. Pierce, N. T., Irber, L., Reiter, T., Brooks, P. & Brown, C. T. Large-scale sequence comparisons with sourmash. *F1000Res* <https://doi.org/10.12688/f1000research.19675.1> (2019).
  101. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
  102. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
  103. Seidl, M. F. et al. Repetitive elements contribute to the diversity and evolution of centromeres in the fungal genus *Verticillium*. *mBio* <https://doi.org/10.1128/mbio.01714-20> (2020).
  104. Shi-Kunne, X. et al. In silico prediction and characterisation of secondary metabolite clusters in the plant pathogenic fungus *Verticillium dahliae*. *FEMS Microbiol. Lett.* **366**, fnz081 (2019).
  105. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
  106. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
  107. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. Tetrascripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).
  108. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
  109. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2010).
  110. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://doi.org/10.48550/arXiv.1303.3997> (2013).
  111. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Giga-science* **10**, giab008 (2021).
  112. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
  113. Bertels, F., Silander, O. K., Pachkov, M., Rainey, P. B. & van Nimwegen, E. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol. Biol. Evol.* **31**, 1077–1088 (2014).
  114. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
  115. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
  116. Katz, L. S. et al. Mashree: a rapid comparison of whole genome sequence files. *J. Open Source Softw.* **4**, 1762 (2019).
  117. Shen, W., Sipos, B. & Zhao, L. SeqKit2: a Swiss army knife for sequence and alignment processing. *iMeta* **3**, e191 (2024).
  118. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
  119. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
  120. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
  121. Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
  122. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
  123. Paysan-Lafosse, T. et al. The Pfam protein families database: embracing AI/ML. *Nucleic Acids Res.* **53**, D523–D534 (2025).
  124. Teufel, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
  125. Dyrlov Bendtsen, J., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
  126. Frandsen, R. J., Andersson, J. A., Kristensen, M. B. & Giese, H. Efficient four fragment cloning for the construction of vectors for targeted gene replacement in filamentous fungi. *BMC Mol. Biol.* **9**, 70 (2008).
  127. Santhanam, P. In *Plant Fungal Pathogens: Methods and Protocols* (eds. Bolton, M. D. & Thomma, B. P. H. J.) (Humana Press, 2012).
  128. de Jonge, R. et al. Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus *Colletotrichum*. *Proc. Natl Acad. Sci. USA* **115**, E5459–E5466 (2018).
  129. Fradin, E. F. et al. Genetic dissection of *Verticillium* wilt resistance mediated by tomato Ve1. *Plant Physiol.* **150**, 320–332 (2009).
  130. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing (2024).
  131. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
  132. Dinno, A. dunn.test: Dunn’s test of multiple comparisons using rank sums. 1.3.6 (2014).
  133. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2008).
  134. Herve, M. RVAideMemoire: testing and plotting procedures for biostatistics. (2025).
  135. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
  136. Hackl, T., Ankenbrand, M. J. & Adrichem, B. van. gggenomes: a grammar of graphics for comparative genomics. (2024).
  137. Guy, L., Roat Kultima, J. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010).
  138. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. GGTREE: an R package for visualization and annotation of phylogenetic trees

with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

139. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
140. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).

## Acknowledgements

Y.S. acknowledges funding through an Overseas Research Fellowship from the Japan Society for the Promotion of Science. B.P.H.J.T. acknowledges funding by the Alexander von Humboldt Foundation in the framework of an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and is furthermore supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2048/1 – Project ID: 390686111. R.B. and M.H. acknowledge funding by the Research Foundation Flanders (FWO project ID: G004022N). We thank Dr. Edgar A. Chavarro-Carrero for providing sequencing read data of the *V. dahliae* genomes assembled in this study.

## Author contributions

Y.S. and B.P.H.J.T. conceived the project. Y.S., M.F.S., and B.P.H.J.T. designed the analyses. R.B. and M.H. contributed to the discovery of the *NLP6* origin. G.C.M.v.d.B. and P.S. performed fungal gene disruption and pathogenicity assays. Y.S., G.C.M.v.d.B., P.S., and B.P.H.J.T. analyzed the data. Y.S. and B.P.H.J.T. wrote the manuscript with feedback from all authors.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-61986-6>.

**Correspondence** and requests for materials should be addressed to Bart P.H.J. Thomma.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025