environmental microbiology

Environmental Microbiology (2018) **20**(4), 1362–1373 doi:10.1111/1462-2920.14037

Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene loss

Xiaoqian Shi-Kunne, Luigi Faino, Grardy C. M. van den Berg, Bart P. H. J. Thomma *† and Michael F. Seidl†
Laboratory of Phytopathology, Wageningen University, Droevendaalsesteeg 1, Wageningen, The Netherlands 6708 PB.

Summary

The fungal genus *Verticillium* contains ten species, some of which are notorious plant pathogens causing vascular wilt diseases in host plants, while others are known as saprophytes and opportunistic plant pathogens. Whereas the genome of *V. dahliae*, the most notorious plant pathogen of the genus, has been well characterized, evolution and speciation of other members of the genus received little attention thus far. Here, we sequenced the genomes of the nine haploid *Verticillium* spp. to study evolutionary trajectories of their divergence from a last common ancestor. Frequent occurrence of chromosomal rearrangement and gene family loss was identified. In addition to ~11 000 genes that are shared at least between two species, only 200–600 species-specific genes occur. Intriguingly, these species-specific genes show different features than the shared genes.

Introduction

Species continuously evolve by genetic variation that enables adaptation to changing and novel environments. In many eukaryotes, such genomic variation is established during sexual reproduction where genetic material of two parents is combined and novel genetic combinations are formed during meiotic recombination (Bell, 1982). Thus, sexual recombination is considered an important driver to establish genetic diversity (Barton and Charlesworth, 1998). However, many species, including fungi, are thought to reproduce strictly asexually (McDonald and Linde, 2002; Heitman *et al*., 2007; Flot *et al*., 2013), and have long been considered limited in their capacity to establish genetic variation. Importantly, even though asexual organisms lack meiotic recombination, adaptive evolution occurs and is established by various mechanisms ranging from single-nucleotide polymorphisms to large-scale structural variations that affect chromosomal shape, organization and gene content (Seidl and Thomma, 2014). Over longer evolutionary time-scales, all these processes establish genetic divergence that may ultimately lead to the emergence of novel species. Typically, these processes can be especially well studied in fungi that have relatively small genomes, which greatly facilitates the establishment of high-quality genome assemblies (Thomma *et al*., 2016).

Previously, gene loss was often neglected as an evolutionary driver, mostly because it was associated with the loss of redundant gene duplicates without apparent functional consequences (Olson, 1999). However, more and more genomic data suggest that gene loss acts as a manifest source of genetic change that may underlie phenotypic diversity (Albalat and Cañestro, 2016). Moreover, reduction or complete loss of gene families has been associated with ecological shifts of fungi (Casadevall, 2008). Human *Malassezia* pathogens that are phylogenetically closest related to plant pathogens such as *Ustilago maydis*, lack fatty acid synthase genes but instead produce secreted lipases to obtain fatty acids from human skin (Xu *et al*., 2007). Gene losses, for instance concerning cell wall-degrading enzymes or secondary metabolite production, have previously also been associated with obligate biotrophic and symbiotic lifestyles of plant-associated fungi (Martin *et al*., 2008; Spanu *et al*., 2010; Duplessis *et al*., 2011).

The fungal genus *Verticillium* consists of ten soil-borne asexual species with different life-styles and host ranges (Inderbitzin *et al*., 2011b; Klosterman *et al*., 2011). Among these, *Verticillium dahliae* is a notorious plant pathogen that causes vascular wilt disease on hundreds of plant species, resulting in large economic losses every year (Fradin and Thomma, 2006;

Received 17 July, 2017; revised 21 December, 2017; accepted 21 December, 2017. *For correspondence. E-mail bart.thomma@wur.nl; Tel. 0031-317-484536; Fax +31-317-418094. †These authors contributed equally to this work.

Klimes *et al*., 2015). Furthermore, also *V. longisporum, V. albo-atrum, V. alfalfae* and *V. nonalfalfae* are plant pathogens, albeit with narrower host ranges (Inderbitzin *et al*., 2011b). The remaining species *V. tricorpus, V. zaregamsianum, V. nubilum, V. isaacii* and *V. klebahnii* are mostly considered saprophytes that thrive on dead organic material and that occasionally cause opportunistic infections (Ebihara *et al*., 2003; Inderbitzin *et al*., 2011b; Gurung *et al*., 2015). Of the ten *Verticillium* species, nine are haploids, while *V. longisporum* is a hybrid that arose from inter-specific hybridisation (Inderbitzin *et al*., 2011a; Depotter *et al*., 2016).

In addition to the genomes of a single strain of *V. alfalfae* and *V. tricorpus* (Klosterman *et al*., 2011; Seidl *et al*., 2015), various strains of *V. dahliae* have been sequenced (Klosterman *et al*., 2011; de Jonge *et al*., 2012). Moreover, for two *V. dahliae* strains a gapless genome assembly has been generated (Faino *et al*., 2015). Comparative genomics between *V. dahliae* strains revealed the occurrence of extensive large-scale genomic rearrangements, likely mediated by erroneous double-stranded break repair, that gave rise to lineage-specific (LS) genomic regions that are enriched for *in planta*-expressed effector genes encoding secreted proteins that mediate host colonization (de Jonge *et al*., 2013; Faino *et al*., 2016). These results raise the question whether pathogenic *Verticillium* spp. evolved by prompting extensive chromosomal rearrangements, thus enabling rapid development of effector gene catalogues that are required to be competitive in arms races with host plants and their immune systems. However, whether these chromosomal rearrangements only occur within pathogenic *Verticillium* spp. remained unknown.

Despite the recent advances in *Verticillium* genomics, the evolutionary history of this genus remains unknown so far. Here, we report high-quality genome assemblies of all haploid *Verticillium* species. We reconstructed ancestral *Verticillium* genomes, and reveal processes of genomic diversification that occurred during *Verticillium* evolution.

Results

High-quality de novo *genome assemblies of the haploid* Verticillium *species*

To infer evolutionary relationships among *Verticillium* spp., we performed comparative genomics with minimum one genome of each of the nine haploid *Verticillium* species. Previously, we sequenced several *V. dahliae* strains as well as a single *V. tricorpus* strain (Klosterman *et al*., 2011; de Jonge *et al*., 2013; Faino *et al*., 2015; Seidl *et al*., 2015). Additionally, the genome of *V. alfalfae* has previously been sequenced (Klosterman *et al*., 2011), as well as the genomes of several *V. nonalfalfae* strains (Bioproject PRJNA283258) (Jelen *et al*., 2016). Here, we sequenced the genomes of a strain of *V. albo-atrum V. isaacii, V. klebahnii, V. nubilum* and *V. zaregamsianum* using the Illumina HiSeq2000 platform. In total, ~18 million paired-end reads (150 bp read length of a 500 bp insert size library) and ~16 million mate-paired reads (150 bp read length of a 5 kb insert size library) were produced per strain. Subsequently, reads were *de novo* assembled into 34–37 Mb draft genomes (Table 1), which is comparable to the assemblies of *V. dahliae*, *V. tricorpus* and *V. alfalfae* (Klosterman *et al*., 2011; de Jonge *et al*., 2013; Faino *et al*., 2015; Seidl *et al*., 2015). All assemblies resulted in less than 100 scaffolds (≥ 1 kb), except for *V. isaacii* strain PD618*, V. nonalfalfae* strain TAB2 and *V. nubilum* strain PD621 that were assembled into 122, 167 and 189 scaffolds respectively (Table 1). Notably, the previously obtained assemblies of *V. tricorpus* strain MUCL9792 and *V. alfalfae* strain VaMs.102 were of significantly lower quality than the assemblies of the newly sequenced *Verticillium* species in this study (Klosterman *et al*., 2011; de Jonge *et al*., 2013; Faino *et al*., 2015; Seidl *et al*., 2015). To obtain better assemblies we decided to sequence an additional strain of each of these two species. Sequencing of *V. tricorpus* strain PD593 and *V. alfalfae* strain PD683 yielded significantly better assembly qualities, with lower numbers of scaffolds (9 for *V. tricorpus* and 14 for *V. alfalfae*) when compared with the previous assemblies (Table 1).

Table 1. Assembly statistics for the various *Verticillium* genomes.

Species	Strain name	Genome size (Mb)	# of Ns per 100 kb	N50 (Mb)	# of contigs (≥0 bp)	# of scaffolds (≥0 bp)	# of scaffolds (≥1000 bp)	CEGMA (%)	BUSCO (%)	# of proteins
V. dahliae	JR2	36.10	0	4.2	8	8	8	95.97	98.75	10 719
V. alfalfae	PD683	32.70	19.36	4.5	40	14	14	94.76	98.82	10 852
V. nonalfalfae	TAB2	34.30	897.53	1.8	793	349	167	95.97	98.47	11 029
V. nubilum	PD621	37.90	9.13	4.7	246	189	189	96.77	99.10	10 550
V. albo-atrum	PD747	36.50	16.26	3.9	34	20	19	94.35	98.96	11 202
V. tricorpus	PD593	35.00	14.52	4.4	71	11	9	95.97	98.82	10 636
V. isaacii	PD618	35.80	62.48	3.1	239	188	122	95.56	98.96	10 798
V. klebahnii	PD401	36.00	35.30	3.2	79	44	37	93.01	99.10	10 998
V. zaregamsianum	PD739	37.10	55.38	3.5	75	46	32	94.35	98.96	11 274

Moreover, the assembly of *V. tricorpus* strain PD593 contained seven scaffolds with telomeric repeats at both ends, suggesting the assembly of seven complete chromosomes (Supporting Information Fig. S1).

To assess completeness of gene space, the assemblies were queried for the presence of orthologous of 248 core eukaryotic gene families using the CEGMA pipeline (Parra *et al.*, 2007). All assemblies contained 93%–97% of these core genes (Table 1). Additionally, we also used the Benchmarking Universal Single-Copy Orthologous (BUSCO) software to assess assembly completeness with 1438 fungal genes as queries (Simão *et al.*, 2015). BUSCO indicated > 98% completeness for each assembly. Considering that we found CEGMA and BUSCO scores of 95.97% and 98.75% respectively, when assessing the gapless genome assembly of *V. dahliae* strain JR2 (Faino *et al.*, 2015), we concluded that a (near) complete gene space was assembled for all *Verticillium* species.

Next, we inferred reference gene annotations by integrating *de novo* and homology-based data using the Maker2 pipeline, making use of 35 predicted fungal proteomes that represent a broad phylogenetic distribution to further guide gene structure annotation (Klosterman *et al.*, 2011; Faino *et al.*, 2015; Seidl *et al.*, 2015). This approach yielded around 11 000 protein-coding genes for each of the genomes, with the highest number of 11 274 genes for *V. zaregamsianum* and the lowest number of 10 636 genes for *V. tricorpus* (Table 1), which is similar to the number of genes identified in previous assemblies of *Verticillium* spp. (Faino *et al.*, 2015; Seidl *et al.*, 2015). However, automatic annotation of *V. dahliae* strain JR2 yielded fewer genes (10 719) than the previously generated annotation (11 430) that, next to RNA-seq data, also involved manual annotation (Faino *et al.*, 2015). In addition to the difference in the number of predicted genes, we also observed differences in genetic features between both annotation methods, such as the overall GC% and the gene, intergenic and intron lengths (Supporting Information Fig. S2). However, for the coding sequences no significant differences in GC% or length were observed (Supporting Information Fig. S2). Thus, besides the number of predicted protein-coding genes, manual annotation mainly increased gene lengths by leveraging UTRs. As all following analyses are based on protein-coding genes, and under-estimation of gene numbers is likely similar for each of the genomes, we used Maker2 annotations for all genome assemblies.

Phylogenetic relationships within the Verticillium *genus*

To better understand the evolutionary events during the evolution of the *Verticillium* genus, determining a robust phylogenetic relationship between *Verticillium* species is crucial. Previously, a phylogeny was constructed that was inferred from a Bayesian analysis of concatenated alignments of four protein-coding marker genes; *actin* (*ACT*), *elongation factor 1-alpha* (*EF*), *glyceraldehyde-3-phosphate dehydrogenase* (*GPD*) and *tryptophan synthase* (*TS*) (Inderbitzin *et al.*, 2011b). To construct the phylogenetic relationships between the nine *Verticillium* species based on whole-genome data, we used concatenated protein sequences of 5228 single-copy orthologous and the out-group species *Sodiomyces alkalinus* to construct a maximum-likelihood phylogeny. The resulting phylogeny reveals two major clades (Fig. 1), which is consistent with the previous analysis (Inderbitzin *et al.*, 2011b). The major clades are the clade Flavexudans (clade A) containing *V. albo-atrum*, *V. isaacii*, *V. klebahnii*, *V. tricorpus* and *V. zaregamsianum*, which are species producing yellow-pigmented hyphae, and the clade Flavnonexudans (clade B) containing *V. alfalfae*, *V. dahliae*, *V. nubilum* and *V. nonalfalfae*, which are species devoid of yellow-pigment.

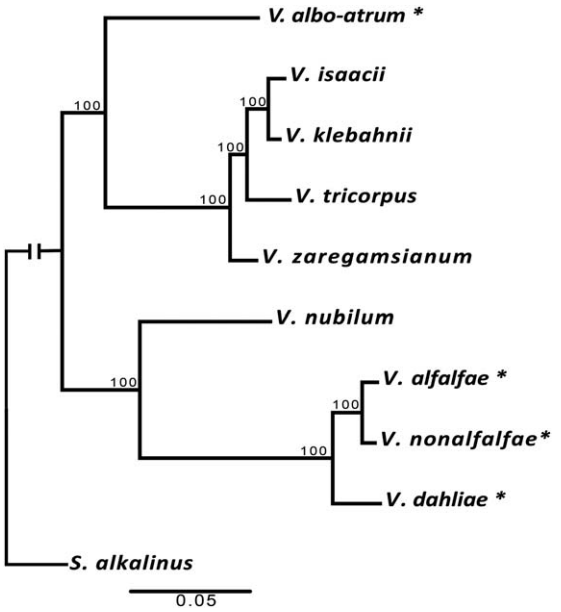

Fig. 1. Phylogenetic tree of *Verticillium* species. Maximum-likelihood phylogeny analysis of *Verticillium* species rooted by *Sodiomyces alkalinus*. The phylogenetic tree is based on 5228 concatenated single-copy orthologous, and the robustness of the tree was assessed using 100 bootstrap replicates. Pathogenic *Verticillium* spp. are marked by asterisks.

Mitochondrial genomes have several unique characteristics, such as a conserved gene content and organization, small size, lack of extensive recombination, maternal inheritance and high mutation rates (Taanman, 1999). This makes them ideal for studying evolutionary relationships among species that diverged during a relatively short period of time. We assembled the complete mitochondrial genomes of each of the *Verticillium* species from paired-end reads using GRAbB (Brankovics *et al.*, 2016), which

Table 2. Mitochondrial genome assemblies of each *Verticillium* strain.

Species	Strain name	Size (bp)	GC content (%)
V. dahliae	JR2	27 178	27.23
V. alfalfae	PD683	25 167	26.84
V. nonalfalfae	TAB2	27 203	26.09
V. nubilum	PD621	27 203	27.35
V. albo-atrum	PD747	28 128	26.58
V. tricorpus	PD593	26 930	27.00
V. isaacii	PD618	26 896	27.10
V. klebahnii	PD401	26 871	27.00
V. zaregamsianum	PD739	26 042	26.78

extracts reads derived from the mitochondrial genomes from the total pool of paired-end sequencing reads using 182 published fungal mitochondrial genome sequences as bait. We assembled all mitochondrial reads per strain into a single circular sequence containing 25–28 Kb with a GC content of 26%–27% (Table 2), in agreement with previous assemblies for *V. dahliae* and *V. nonalfalfae* (Klosterman *et al.*, 2011; Jelen *et al.*, 2016). We subsequently annotated each of the mitochondrial genomes, identifying 15 protein-coding genes. For each mitochondrial genome, all genes were encoded on a single strand and in the same direction (Jelen *et al.*, 2016) (Fig. 2B). The whole-mitochondrial-genome alignments were used to construct a maximum likelihood phylogeny (Fig. 2A), revealing the same topology as the phylogenetic tree that is based on the nuclear genome (Fig. 1). Thus, the mitochondrial genome-based phylogenetic tree further supports the robustness of the *Verticillium* whole-genome phylogeny based on the nuclear genome assemblies, and further corroborates previous phylogenies derived by a limited set of marker genes (Inderbitzin *et al.*, 2011b).

It was reported recently that the mitochondrial genomes of both *V. nonalfalfae* and *V. dahliae* carry a *Verticillium*-specific region with a length of 789 bp, named *orf414* (Jelen *et al.*, 2016). When we aligned the mitochondrial genomes of each of the *Verticillium* species and the close relative *S. alkalinus*, we found that *orf414* is absent from *S. alkalinus* (Fig. 2A). However, *orf414* is not conserved in the mitochondrial genomes of all *Verticillium* species, as it is only found in *V. nonalfalfae*, *V. dahliae* and *V. nubilum*. Thus, *orf414* cannot be used as a *Verticillium*-specific diagnostic marker (Jelen *et al.*, 2016).

Reconstruction of ancestor genomes

The degree of synteny between extant species reveals essential information about the evolution of species from their last common ancestor (Hane *et al.*, 2011; Lv *et al.*, 2011). Based on the length of the sequence stretch showing a relatively high degree of sequence similarity and on the degree of co-linearity, synteny can be defined as macro-, micro- or mesosynteny (Hane *et al.*, 2011). Macrosynteny occurs when large ($>$ 20 kb) blocks of genes are shared between species in the same order and orientation, while microsynteny involves smaller ($<$ 20 kb) segments with relatively few genes. Like microsynteny, mesosynteny involves relatively small ($<$ 20 kb) segments, but whereas microsynteny implies that genes are conserved with the same order and orientation, mesosynteny indicates that genes within the segment are conserved with absence of co-linearity. To determine the type of synteny between the extant *Verticillium* species, we performed pairwise alignments of the genomes of the two most divergent *Verticillium* species, *V. dahliae* and *V. albo-atrum* (Hane

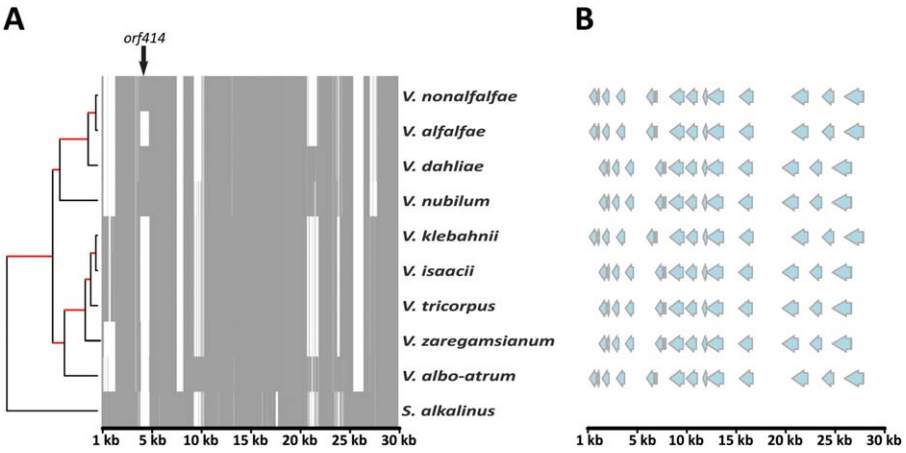

Fig. 2. Mitochondrial genome alignments.
A. Whole mitochondrial genome alignment of all *Verticillium* species and *S. alkalinus*. Grey and white colours represent presence and absence of genomic regions respectively. The whole mitochondrial genome alignments were used for constructing a maximum likelihood phylogenetic tree. The robustness of the topology was assessed using 100 bootstrap replicate (branches with maximum bootstrap values are in red).
B. Graphic presentation of positions of mitochondrial protein-coding genes and their orders. [Colour figure can be viewed at wileyonlinelibrary.com]

et al., 2011). These pairwise alignments showed clear regions of macro- and microsynteny, but no mesosynteny (Supporting Information Fig. S3). Subsequently, we performed similar analyses for the other *Verticillium* species, confirming the occurrence of macro- and microsynteny, and the absence of mesosynteny between *Verticillium* spp. (Supporting Information Fig. S4).

Considering that extensive macrosynteny occurs between the extant *Verticillium* spp. we aimed to reconstruct the evolution of the *Verticillium* genus from its last common ancestor. Inferring the genome organization and gene content of ancestral species has the potential to provide detailed information about the recent evolution of descendant species. However, reconstruction of ancestral genome architectures, followed by integrating into evolutionary frameworks, has only been achieved for a limited number of species (Nakatani *et al.*, 2007; Gordon *et al.*, 2009; Vakirlis *et al.*, 2016). This is largely due to either the unavailability of sequences from multiple closely related species, or due to the high fragmentation of the genome assemblies used. To minimize the influence of fragmented draft genome assemblies, we only considered the largest scaffolds that comprise 95% of the total set of protein-coding genes for each of the genomes (Supporting Information Table S1), and constructed ancestral genome organizations that preceded each speciation event using SynChro (Drillon *et al.*, 2014) and AnChro (Vakirlis *et al.*, 2016). SynChro identifies conserved synteny blocks between pairwise comparisons of extant genomes, after which AnChro infers the ancestral gene order by comparing these synteny blocks. To validate the accuracy of AnChro, we first reconstructed the genome of the last common ancestor of *V. dahliae* and *V. alfalfae*. We did this separately for two complete and gapless genome assemblies of the extant *V. dahliae* strains JR2 and VdLs17 that, despite extensive genomic rearrangements and the presence of LS sequences (de Jonge *et al.*, 2012; de Jonge *et al.*, 2013; Faino *et al.*, 2016), each contain eight chromosomes (Faino *et al.*, 2015). Irrespective whether the genome of strain JR2 or VdLs17 was used, the resulting ancestor has nine scaffolds that are similarly organized. To compare the two ancestors, synteny blocks were constructed and aligned using SynChro, revealing an overall identical genome structure with only a few genes that lack homologue (Supporting Information Fig. S5). Owing to the different lineage-specific regions in the JR2 and VdLs17 genomes, the number of genes in the ancestor varies slightly, with 9165 and 9250 protein-coding genes based on the genome of JR2 or VdLs17 respectively. Thus, we concluded that the AnChro software is suitable for reconstruction of ancestral genomes in the genus *Verticillium*.

Ancestral genomes were reconstructed for all the nodes in the phylogenetic tree, resulting in less than 20 scaffolds at each individual node (Fig. 3). Using SynChro to determine synteny blocks between the last common ancestor and the ancestor-derived genomes that served as input for ReChro (Vakirlis *et al.*, 2016), the number of rearrangements that occurred in each branch of the tree was determined. In total, 498 rearrangements, including chromosomal fusions and fissions, occurred during the evolution from the last common ancestor to the nine extant haploid *Verticillium* species (Fig. 3). Yet, considerable variation occurred between species ranging from 69 rearrangements for *V. tricorpus* to up to 205 for *V. alboatrum* (Fig. 3).

To assess whether genomic rearrangements occurred in a clock-like fashion during the evolution of *Verticillium* species, we related the number of rearrangements per branch with the evolutionary time approximated by the branch lengths inferred from the phylogenetic tree (Fig. 1). The branch lengths were represented by either the numbers of substitutions per site (Fig. 1), or the relative divergence time that was estimated based on artificial dating of the last common ancestor of *Verticillium* and *S. alkalinus* to 100 units of time (Supporting Information Fig. S6). The number of rearrangements per branch showed significant correlation with both representations of branch length ($R^2 = 0.3679$, $P = 0.007541$ for substitution per site and $R^2 = 0.7623$, $P = 6.174$ e–06 for relative divergence time) (Supporting Information Fig. S7).

Determination of gene family expansions and contractions

To monitor gene family changes during evolution, we estimated gene family size expansion (gains) and contraction (losses) on each branch using CAFE (Han *et al.*, 2013). In this analysis, we considered all gene families that are present in at least two *Verticillium* species. CAFE models the evolution of gene family size across a species phylogeny under a birth–death model of gene gain and loss and simultaneously reconstructs ancestral gene family sizes for all internal nodes, allowing the detection of expanded or contracted families within lineages. The analysis revealed that the last common *Verticillium* ancestor contained 11 902 families with 12 631 genes. Intriguingly, *Verticillium* species generally underwent more extensive gene losses than gains (Fig. 4). This pattern of more extensive gene losses than gains was further confirmed by inferring gene family gains and losses by reconciliation of 10 071 gene trees with established *Verticillium* species phylogeny (Fig. 1; Supporting Information Fig. S8). To obtain functional descriptions of gained and lost gene families, we searched for the corresponding Cluster of Orthologous Group (COG) functional categories in the eggNOG database (Huerta-Cepas *et al.*, 2015). In both the A and B clade, the most prevalent COG functional category among gained and lost gene families is ‘carbohydrate metabolism and

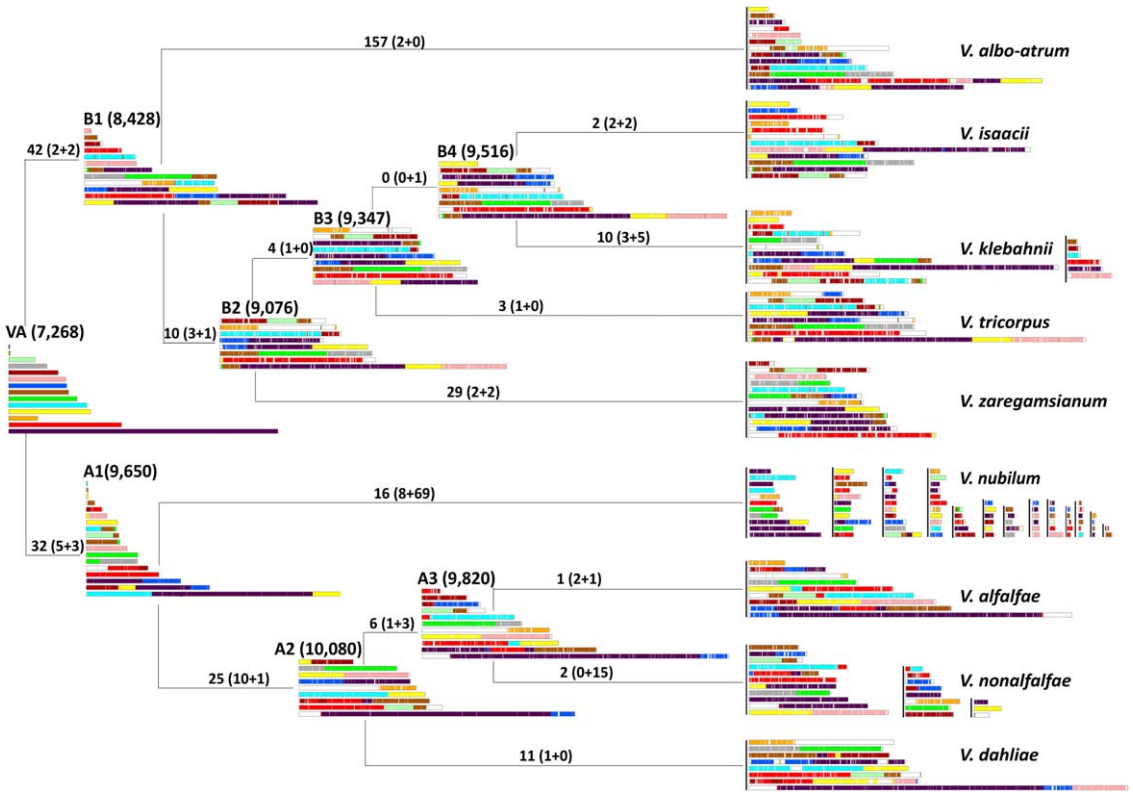

Fig. 3. Chromosomal history of *Verticillium* genomes. Genome structure changes from the most common ancestor (VA) to the nine extant *Verticillium* species. The number of genes of each ancestor is indicated in brackets above the chromosomes. The number of chromosome rearrangements, i.e., the sum of translocations and inversions is indicated above each branch. The number of fusions and fissions between two genomes respectively, are indicated in brackets. [Colour figure can be viewed at wileyonlinelibrary.com]

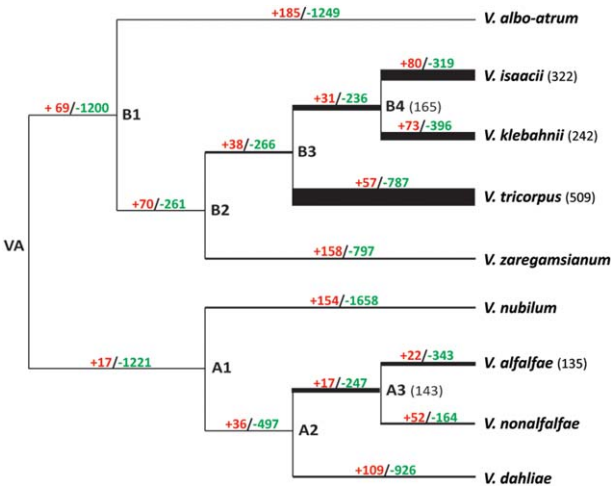

Fig. 4. Evolution of *Verticillium* gene repertoire. The number of expanded (in red) and contracted (in green) gene families was estimated on each branch of the tree under a birth–death evolutionary model. The thickened branches represent the abundance of gene families that evolved rapidly ($P < 0.05$) and the exact number of gene families is indicated after each node name. [Colour figure can be viewed at wileyonlinelibrary.com]

transport' (Supporting Information Fig. S9). To obtain more detailed functional predictions of the gene families that underwent gains and losses, we first annotated Pfam domains for each gene family (Supporting Information Data S1 and Data S3). Subsequently, using a Hypergeometric test with a false discovery rate (FDR)-corrected *P* value of < 0.05, we identified nine and 14 Pfam domains that are enriched in the gained and lost gene families respectively (Supporting Information Table S2). Among the enriched Pfam domains of clade B, two belong to carbohydrate enzyme gene families, GH3 and GH35. The remainder of the Pfam domains relates to transporter activities, metabolic processes or unknown functions. The enriched Pfam domains of the lost gene families show more diverse functions, including transcription factors and cytochrome P450 activity. Among all expanded and contracted genes families, we found 1081 gene families that are significantly more variable in size during evolution ($P < 0.05$). Subsequently, we estimated in which branches these gene families evolved more rapidly using the Viterbi algorithm of CAFE. This algorithm calculates a *P* value for each of these 1081 gene families on each of the branches of the phylogenetic tree. A *P* value that is below the

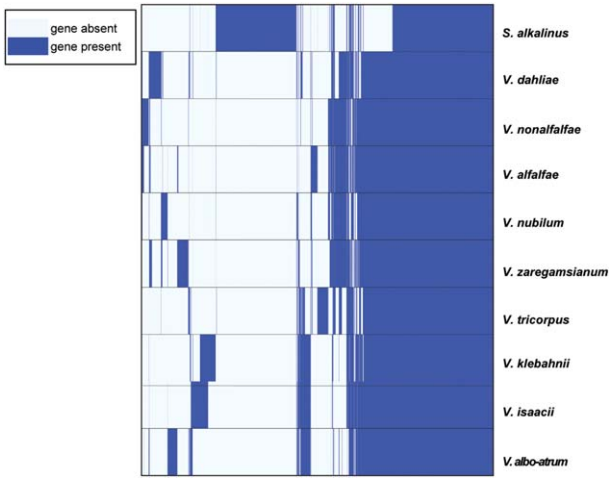

Fig. 5. *Verticillium* gene family conservation. Presence or absence of gene families among *Verticillium* species are indicated by dark and light blue respectively. The gene families (columns) were ordered by hierarchical clustering. [Colour figure can be viewed at wileyonlinelibrary.com]

threshold ($P < 0.05$) indicates that the respective gene family evolved more rapidly on that particular branch. This analysis revealed that the branches between *V. tricorpus, V. isaacii, V. klebahnii* and their last common ancestor (B4) evolved most rapidly (involving 509, 322, 242 and 165 gene families respectively), followed by the branches of *V. alfalfae* and *V. nonalfalfae* with their last common ancestor (A3) (143 and 135 for *V. alfalfae* and A3 respectively) (Fig. 4). Next, the overrepresentation of Pfam domains in the most rapidly evolved gene families was assessed using a Hypergeometric test with a FDR-corrected *P* value of < 0.05. Under these conditions, only a single Pfam domain PF01636 was found to be enriched ($P = 1$ e–06), concerning a phosphotransferase enzyme family (APH) that is predicted to consist of antibiotic resistance proteins (Sarwar and Akhtar, 1990; Trower and Clark, 1990; Nurizzo *et al*., 2003). This Pfam domain was also found to be enriched in the lost gene families in *V. isaacii* (Supporting Information Table S2).

Although we observed considerable gene losses and gains along the different branches of the tree, we also observed a large number of shared genes among all extant species. In total, 7538 genes are common among all *Verticillium* spp., of which 1689 are absent from *S. alkalinus* (Fig. 5). Interestingly, only as little as 123 genes are clade A-specific and 288 genes clade B-specific. These genes underwent losses in the reciprocal clade in a single event only at the last common ancestor of the respective clades, rather than multiple independent losses. For each genome, less than 5% of the total protein-coding genes are species-specific (Fig. 5; Supporting Information Data S2). These species-specific genes were neither enriched for any Pfam domain nor for genes that encode secreted proteins. To infer the evolutionary trajectory of these species-specific genes, we searched for homologues of these genes in the proteomes of 383 fungal species. Up to 50% of these species-specific genes had homologues in non-*Verticillium* species, suggesting that the species-specific occurrence in *Verticillium* is the result from losses from other *Verticillium* species during evolution (Supporting Information Table S3).

To assess the characteristics of the species-specific genes versus shared genes, we compared features such as GC content, gene lengths, intergenic lengths and intron lengths. Interestingly, when compared with core genes, species-specific genes significantly differ in GC content of gene sequences as well as of coding sequences, and in gene length (Fig. 6A–C). Some species also show significant differences in intron lengths and intergenic lengths between species-specific and core genes (Figure 6D-E). Intriguingly, most of the species-specific genes of *V. dahliae* strain JR2 are not, or only lowly, expressed *in vitro* (Fig. 6F).

Discussion

In this study, we investigated genomic changes that occurred during evolution of the *Verticillium* genus from the last common ancestor to the currently recognized extant species. During evolution, two or more independent populations within a single species may diverge and slowly start to separate, which may ultimately lead to reproductive isolation and thus to speciation. Initially the separated species will have chromosomes that share the gene content (synteny) as well as the structure and order (co-linearity). Over time, the degree of synteny and co-linearity will degrade through various processes, including chromosomal rearrangements, segmental duplications, gene losses and gene gains until, ultimately, orthologous genes in one species occur randomly in the genome of the other.

We previously noted an unexpectedly high number of chromosomal rearrangements between strains of *V. dahliae* (de Jonge *et al*., 2013; Faino *et al*., 2015; Faino *et al*., 2016). We have speculated that the extent of rearrangements may be associated with the fact that, despite being asexual, *V. dahliae* is a successful broad host-range pathogen, reasoning that it would permit for the rapid adaptations that are required to be compatible in the arms race with host immune systems (Seidl and Thomma, 2014; Faino *et al*., 2016). From this hypothesis it would follow that the other species, being much less ubiquitous and successful pathogens, would not experience such drastic genomic rearrangements. However, our genomic reconstructions revealed that large-scale genomic rearrangements frequently occurred during evolution of the *Verticillium* genus. Moreover, our data seems to suggest that rearrangements occurred even more frequent in clade

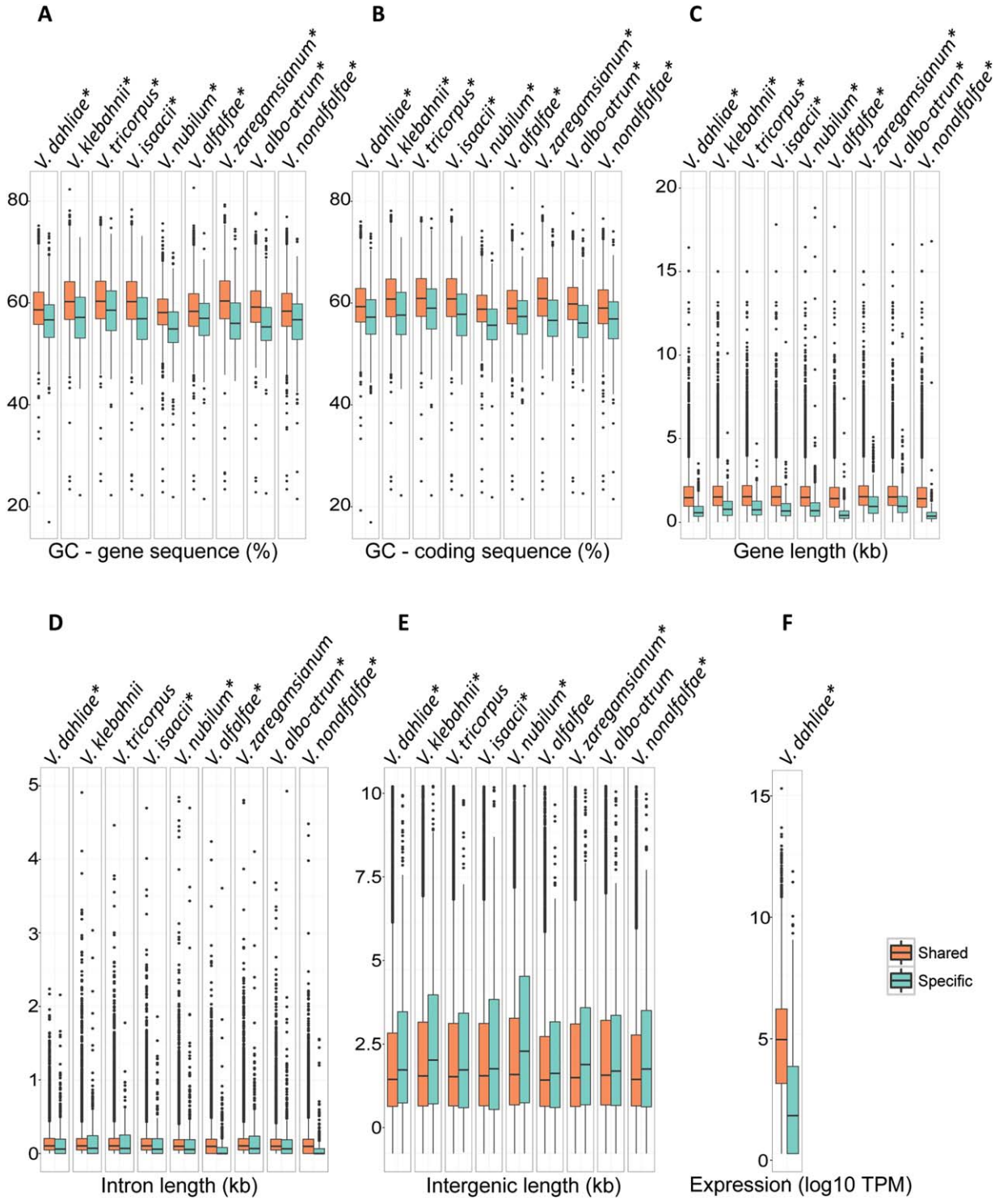

Fig. 6. Gene feature comparisons between specific and shared genes. The Wilcoxon rank sum test ($P < 0.01$, indicated by asterisk) was used to detect significant differences between species-specific and shared genes of each species.
A. GC content based on gene sequences.
B. GC content based on coding sequences.
C. Intron length.
D. Intergenic length.
E. *V. dahliae in vitro* expression. [Colour figure can be viewed at wileyonlinelibrary.com]

B that mostly harbours non-pathogenic species. This observation makes it unlikely that the occurrence of the rearrangements themselves is a major contributor to the pathogenicity of *V. dahliae*. Previously, we have shown that the highly variable LS regions of *V. dahliae* evolved by genomic rearrangements (de Jonge *et al*., 2013; Faino *et al*., 2016). We also showed that LS regions that harbour *in planta*-expressed effector genes arose by segmental duplications, likely generating the genetic material that subsequently has the freedom to diverge into novel functions (de Jonge *et al*., 2013; Faino *et al*., 2016). Thus, LS regions play important roles in adaption of *V. dahliae*. However, whether such LS regions also occur in other *Verticillium* spp. still needs to be investigated.

Besides extensive genomic rearrangements, conspicuous gene losses occurred during *Verticillium* evolution. This suggests that gene losses contribute to genetic divergence across the various *Verticillium* lineages. The impact of gene loss on evolution and pathogenicity is perhaps most evident for obligate biotrophic plant pathogens whose growth and reproduction entirely depends on living plant cells, such as the powdery mildews (Spanu *et al*., 2010). Their genomes show massive genome-size expansion, owing to retrotransposon proliferation, and extensive gene losses concerning enzymes of primary and secondary metabolism that are responsible for loss of autotrophy and dependence on host plants in an exclusively biotrophic lifestyle (Spanu *et al*., 2010). Recently, the evolutionary history of lineages of grass powdery mildew fungi that have a rather peculiar taxonomy with only one described species (*Blumeria graminis*) that is separated into various *formae speciales* (*ff. spp.*) was reconstructed. It was observed that different processes shaped the diversification of *B. graminis*, including co-evolution with the host species for some of the *ff. spp.*, host jumps, host range expansions, lateral gene flow and fast radiation (Menardo *et al*., 2017). Arguably, as highly adapted and obligate biotrophic pathogens, powdery mildews underwent many host species–specific adaptations, which are perhaps not or less required for saprophytes and facultative and broad host range pathogens such as *Verticillium* species. Among the lost gene families in *Verticillium* species, we found that enriched Pfam domains relate to various functions, suggesting that the lost gene families of *Verticillium* spp. do not revolved around a particular function.

Functions of species-specific genes are often found to be associated to species-specific adaptations to a certain environment (Domazet-Loso and Tautz, 2003). For example, morphological and innate immune differences among *Hydra* spp. are controlled by species-specific genes (Khalturin *et al*., 2008; Khalturin *et al*., 2009). In plant pathogens, many effectors characterized so far are species-specific and facilitate virulence on a particular host plant (de Jonge *et al*., 2011) However, these species-specific effectors are frequently mutated, or even purged, to overcome host recognition (Cook *et al*., 2015). Moreover, it has been shown that compared to shared genes, species-specific genes are frequently shorter (Lipman *et al*., 2002), evolving quicker and less expressed (Domazet-Loso and Tautz, 2003; Plissonneau *et al*., 2016). Consistent with this observation, our results show those species-specific genes differ in their characteristics when compared with core genes. It has often been claimed that these distinct characteristics are hallmarks of gene structure degeneration and that these species-specific genes are more likely to go extinct (Palmieri *et al*., 2014; Plissonneau *et al*., 2016).

Experimental procedures

Genome sequencing and assembly

DNA was isolated from mycelium of 3-day-old cultures grown in potato dextrose broth (PDB) at 28°C as described previously (Seidl *et al*., 2015). Of each strain, two libraries (500 bp and 5 Kb insert size) were sequenced using the Illumina High-throughput sequencing platform (KeyGene N.V., Wageningen, The Netherlands). Genome assemblies were performed using the A5 pipeline (Tritt *et al*., 2012), and sequence gaps were filled using SOAP GapCloser (Luo *et al*., 2012). Next, QUAST (Gurevich *et al*., 2013) was used to calculate genome statistics. Illumina sequence reads and assemblies were deposited in NCBI (Bioproject PRJNA392396).

Gene prediction and annotation

Protein-coding genes were *de novo* annotated with the Maker2 pipeline (Holt and Yandell, 2011) using 35 predicted fungal proteomes and the previously annotated proteomes of *V. dahliae* and *V. tricorpus* to guide gene structure annotation (Faino *et al*., 2015; Seidl *et al*., 2015). Secretome prediction was described previously (Seidl *et al*., 2015).

Orthologous analysis and tree building

Orthologous groups were determined using OrthoMCL (Li *et al*., 2003). The species phylogenetic tree was generated using 5228 single-copy orthologous that are conserved among all of the genomes. Individual families were aligned using mafft (LINSi; v7.04b) (Katoh *et al*., 2002) and subsequently concatenated. Maximum likelihood phylogeny was inferred using RAxML (v8.2.4) with the GAMMA model of rate heterogeneity and the Whelan and Goldman (WAG) model of amino acid substitutions (Stamatakis, 2014). The robustness of the inferred phylogeny was assessed by 100 rapid bootstrap approximations.

Mitochondrial genome assembly and comparison

Sequencing reads of mitochondrial genomes of each *Verticillium* species were extracted from raw paired-end reads using GRAbB using 182 already published fungal mitochondrial genome sequences as bait (Brankovics *et al*., 2016).

Subsequently, GRAbB assembled all the mitochondrial reads into single circular sequences. To determine the mitochondrial protein-coding genes of each species, we queried the already published *Verticillium dahliae* mitochondrial protein sequences using BLAST (tblastn) against each mitochondrial genome. Whole mitochondrial genome alignment was performed with mafft (default setting) (Katoh *et al*., 2002), and the likelihood phylogenetic tree was built using RAxML (v7.6.3) with the GAMMA model of rate heterogeneity and the GTR model of nucleotide substitutions (Stamatakis, 2014). The robustness of the inferred phylogeny was assessed by 100 rapid bootstrap approximations.

Ancestor genome reconstruction and comparison

Ancestral genomes were constructed using CHROnicle package that comprises SynChro (Drillon *et al*., 2014), ReChro and Anchro (Vakirlis *et al*., 2016). Conserved synteny blocks were identified between pairwise combinations of genomes with SynChro (Drillon *et al*., 2014). The synteny block stringency delta, which determines the maximum number of intervening Reciprocal Best Hits (RBH) allowed between anchors within a synteny block was set to three. Subsequently, ReChro was used for estimating genome rearrangements from ancestral to extant genomes.

Gene family gains and losses

Gene family gain/loss analysis by inferring ancestral gene number counts was carried out using CAFE (Han *et al*., 2013). Additionally, we inferred gene family gains and losses by reconciliation of gene trees. The sequences of the gene families ($>$ 3 members) were aligned using mafft with default settings (Katoh *et al*., 2002). We constructed phylogenetic trees for each gene family using RAxML (v8.2.4) with the GAMMA model of rate heterogeneity and the Whelan and Goldman (WAG) model of amino acid substitutions (Stamatakis, 2014). Subsequently, we reconciled the gene family trees with the *Verticillium* species tree using NOTUNG (Chen *et al*., 2000). The trees were reconciled using a cost of 1.5 for a duplication event and 1 for a loss event. The trees were rooted so that the number of duplication and loss events is minimized. Cluster of Orthologous Groups (COG) functional categories were predicted using eggNOG (Huerta-Cepas *et al*., 2015). Pfam function domains were predicted using InterProScan (Jones *et al*., 2014). For analyses of Pfam domain associated with gene families (OrthoMCL clusters), Pfam domains were assigned to specific families (clusters) only when present in at least half of the homologues within each gene family. Pfam enrichment of gene families of interests was carried out using hypergeometric tests, and significance values were corrected using the Benjamini–Hochberg false discovery method (Benjamini and Hochberg, 1995).

Acknowledgement

We thank Drs Inderbitzin and Subbarao for sharing fungal isolates and Sander Rodenburg for help with scripting.

Conflict of Interest

The authors declare no competing interests.

References

Albalat, R., and Cañestro, C. (2016) Evolution by gene loss. *Nat Rev Genet* **17:** 379–391.

Barton, N.H., and Charlesworth, B. (1998) Why sex and recombination?. *Science* **281:** 1986–1990.

Bell, G. (1982) *The Masterpiece of Nature: The Evolution and Genetics of Sexuality.* Berkley: University California Press.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* **57:** 289–300.

Brankovics, B., Zhang, H., van Diepeningen, A.D., van der Lee, T.A.J., Waalwijk, C., de Hoog, G.S., and Gardner, P.P. (2016) GRAbB: selective assembly of genomic regions, a new niche for genomic research. *PLoS Comp Biol* **12:** e1004753.

Casadevall, A. (2008) Evolution of intracellular pathogens. *Annu Rev Microbiol* **62:** 19–33.

Chen, K., Durand, D., and Farach-Colton, M. (2000) NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol* **7:** 429–447.

Cook, D.E., Mesarich, C.H., and Thomma, B.P. (2015) Understanding plant immunity as a surveillance system to detect invasion. *Annu Rev Phytopathol* **53:** 541–563.

de Jonge, R., Bolton, M.D., and Thomma, B.P. (2011) How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr Opin Plant Biol* **14:** 400–406.

de Jonge, R., Peter van Esse, H., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., *et al*. (2012) Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc Natl Acad Sci USA* **109:** 5110–5115.

de Jonge, R., Bolton, M.D., Kombrink, A., van den Berg, G.C., Yadeta, K.A., and Thomma, B.P. (2013) Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res* **23:** 1271–1282.

Depotter, J.R., Seidl, M.F., Wood, T.A., and Thomma, B.P. (2016) Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. *Curr Opin Microbiol* **32:** 7–13.

Domazet-Loso, T., and Tautz, D. (2003) An evolutionary analysis of orphan genes in Drosophila. *Genome Res* **13:** 2213–2219.

Drillon, G., Carbone, A., Fischer, G., and Fairhead, C. (2014) SynChro: a fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. *PLoS One* **9:** e92621.

Duplessis, S., Cuomo, C.A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., *et al*. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci USA* **108:** 9166–9171.

Ebihara, Y., Uematsu, S., Nagao, H., Moriwaki, J., and Kimishima, E. (2003) First report of *Verticillium tricorpus* isolated from potato tubers in Japan. *Mycoscience* **44:** 481–488.

Faino, L., Seidl, M.F., Datema, E., van den Berg, G.C., Janssen, A., Wittenberg, A.H., and Thomma, B.P. (2015) Single-molecule real-time sequencing combined with optical

mapping yields completely finished fungal genome. *mBio* **6:** e00936–e00915.

Faino, L., Seidl, M.F., Shi-Kunne, X., Pauper, M., van den Berg, G.C., Wittenberg, A.H., and Thomma, B.P. (2016) Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res* **26:** 1091–1100.

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E.G., *et al.* (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* **500:** 453–457.

Fradin, E.F., and Thomma, B.P. (2006) Physiology and molecular aspects of Verticillium wilt diseases caused by *V. dahliae* and *V. albo-atrum*. *Mol Plant Pathol* **7:** 71–86.

Gordon, J.L., Byrne, K.P., Wolfe, K.H., and Zhang, J. (2009) Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet* **5:** e1000485.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29:** 1072–1075.

Gurung, S., Short, D.P., Hu, X., Sandoya, G.V., Hayes, R.J., Koike, S.T., and Subbarao, K.V. (2015) Host range of *Verticillium isaacii* and *Verticillium klebahnii* from artichoke, spinach, and lettuce. *Plant Dis* **99:** 933–938.

Han, M.V., Thomas, G.W., Lugo-Martinez, J., and Hahn, M.W. (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* **30:** 1987–1997.

Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H., Goodwin, S.B., and Oliver, R.P. (2011) A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol* **12:** R45.

Heitman, J., Kronstad, J.W., Taylor, J., and Casselton, L. (2007) *Sex in Fungi: Molecular determination and evolutionary implications*. Washington, DC: ASM Press.

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

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., *et al.* (2015) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44:** D286–D293.

Inderbitzin, P., Davis, R.M., Bostock, R.M., and Subbarao, K.V. (2011a) The ascomycete *Verticillium longisporum* is a hybrid and a plant pathogen with an expanded host range. *PLoS One* **6:** e18260.

Inderbitzin, P., Bostock, R.M., Davis, R.M., Usami, T., Platt, H.W., and Subbarao, K.V. (2011b) Phylogenetics and taxonomy of the fungal vascular wilt pathogen Verticillium, with the descriptions of five new species. *PLoS One* **6:** e28341.

Jelen, V., de Jonge, R., Van de Peer, Y., Javornik, B., Jakše, J., and Nowrousian, M. (2016) Complete mitochondrial genome of the Verticillium-wilt causing plant pathogen *Verticillium nonalfalfae*. *PLoS One* **11:** e0148525.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30:** 1236–1240.

Katoh, K., Misawa, K., Kuma, K.I., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30:** 3059–3066.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., and Bosch, T.C. (2009) More than just orphans: are taxonomically-restricted genes important in evolution?. *Trends Genet* **25:** 404–413.

Khalturin, K., Anton-Erxleben, F., Sassmann, S., Wittlieb, J., Hemmrich, G., Bosch, T.C.G., and Patel, N. (2008) A novel gene family controls species-specific morphological traits in Hydra. *PLoS Biol* **6:** e278.

Klimes, A., Dobinson, K.F., Thomma, B.P., and Klosterman, S.J. (2015) Genomics spurs rapid advances in our understanding of the biology of vascular wilt pathogens in the genus *Verticillium*. *Annu Rev Phytopathol* **53:** 181–198.

Klosterman, S.J., Subbarao, K.V., Kang, S., Veronese, P., Gold, S.E., Thomma, B.P.H.J., *et al.* (2011) Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Patholog* **7:** e1002137.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13:** 2178–2189.

Lipman, D.J., Souvorov, A., Koonin, E.V., Panchenko, A.R., and Tatusova, T.A. (2002) The relationship of protein conservation and sequence length. *BMC Evol Biol* **2:** 20.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1:** 18.

Lv, J., Havlak, P., and Putnam, N.H. (2011) Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes. *BMC Bioinform* **12:** S11.

Martin, F., Aerts, A., Ahren, D., Brun, A., Danchin, E.G.J., Duchaussoy, F., *et al.* (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452:** 88–92.

McDonald, B.A., and Linde, C. (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol* **40:** 349–379.

Menardo, F., Wicker, T., and Keller, B. (2017) Reconstructing the evolutionary history of powdery mildew lineages *(Blumeria graminis*) at different evolutionary time scales with NGS data. *Genome Biol Evol* **9:** 446.

Nakatani, Y., Takeda, H., Kohara, Y., and Morishita, S. (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* **17:** 1254–1265.

Nurizzo, D., Shewry, S.C., Perlin, M.H., Brown, S.A., Dholakia, J.N., Fuchs, R.L., *et al.* (2003) The crystal structure of aminoglycoside-3′-phosphotransferase-IIa, an enzyme responsible for antibiotic resistance. *J Mol Biol* **327:** 491–506.

Olson, M.V. (1999) When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet* **64:** 18–23.

Palmieri, N., Kosiol, C., and Schlötterer, C. (2014) The life cycle of Drosophila orphan genes. *eLife* **3:** e01311.

Parra, G., Bradnam, K., and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23:** 1061–1067.

Plissonneau, C., Stürchler, A., and Croll, D. (2016) The evolution of orphan regions in genomes of a fungal pathogen of wheat. *mBio* **7:** e01231–e01216.

Sarwar, M., and Akhtar, M. (1990) Cloning of aminoglycoside phosphotransferase (APH) gene from antibiotic-producing

strain of Bacillus circulans into a high-expression vector, pKK223-3. Purification, properties and location of the enzyme. *Biochem J* **268:** 671–677.
Seidl, M.F., and Thomma, B.P. (2014) Sex or no sex: evolutionary adaptation occurs regardless. *Bioessays* **36:** 335–345.
Seidl, M.F., Faino, L., Shi-Kunne, X., van den Berg, G.C., Bolton, M.D., and Thomma, B.P. (2015) The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. *Mol Plant-Microbe Interact* **28:** 362–373.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31:** 3210–3212.
Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., *et al*. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330:** 1543–1546.
Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.
Taanman, J.-W. (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta Bioenerg* **1410:** 103–123.
Thomma, B.P., Seidl, M.F., Shi-Kunne, X., Cook, D.E., Bolton, M.D., van Kan, J.A., and Faino, L. (2016) Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genet Biol* **90:** 24–30.
Tritt, A., Eisen, J.A., Facciotti, M.T., and Darling, A.E. (2012) An integrated pipeline for de novo assembly of microbial genomes. *PLoS One* **7:** e42304.
Trower, M.K., and Clark, K.G. (1990) PCR cloning of a streptomycin phosphotransferase (aphE) gene from Streptomyces griseus ATCC 12475. *Nucleic Acids Res* **18:** 4615.
Vakirlis, N., Sarilar, V., Drillon, G., Fleiss, A., Agier, N., Meyniel, J.-P., *et al*. (2016) Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Res* **26:** 918–932.
Xu, J., Saunders, C.W., Hu, P., Grant, R.A., Boekhout, T., Kuramae, E.E., *et al*. (2007) Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci USA* **104:** 18730–18735.

Supporting information

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

Fig. S1. Overview of the draft genome assembly of *V. tricorpus*, strain PD593. Schematic representation of the eight largest scaffolds in the genome assembly of *V. tricorpus* strain PD593. Characteristic fungal telomeric repeats are displayed on the ends of the scaffolds (indicated by red colour).
Fig. S2. Gene feature comparisons between automatically and manually annotated genes of *V dahliae* strain JR2. A. GC content. B. Gene/coding sequence length. C. Intergenic length. D. Intron length.
Fig. S3. Dot-plot comparison between *V. dahliae* and *V. albo-atrum*. The sixframe translations of both genomes were compared using Promer (Mummer 3.0). Homologous regions are plotted as dots and are colour coded for percentage similarity. Macro- or microsynteny is indicated by long or short diagonal lines with a slope that is positive or negative depending on whether the genes align in the same or inverted order.
Fig. S4. A–D. Pairwise dot-plot comparison between nine haploid *Verticillium* species. Macro- or microsynteny is indicated by long or short diagonal lines with a slope that is positive or negative depending on whether the genes align in the same or inverted order.
Fig. S5. Chromosome alignments between two ancestor genomes. Two ancestor genomes were reconstructed based on the genome of two *V. dahliae* strain JR2 and VdLs17 respectively (indicated by red and blue dots respectively). Green and red plain lines highlight homology relationships with same and inverted directions respectively.
Fig. S6. Ultrametric phylogeny of *Verticillium* species and *S. alkalinus*. The ultrametric tree derived by a maximum likelihood analysis of concatenated single-copy orthologous (Fig. 1). Averaged divergence times per branch are reported relative to an arbitrary age of the last common ancestor of *Verticillium* spp. and *S. alkalinus* (set at 100).
Fig. S7. Correlations between the branch lengths and number of rearrangements. A. Branch length is represented by number of substitutions per site (Fig. 1), B. Branch length is represented by evolutionary distance (Supporting Information Fig. S3)
Fig. S8. Evolution of *Verticillium* gene repertoire. Numbers of expanded (in red) and contracted (in green) gene families were inferred by a tree reconciliation approach.
Fig. S9. Percentages for each COG category of gained or lost gene families. **A** Clade A, **B** Clade B.
Table S1. Minimum number of scaffolds that comprises at least 95% of total proteomes of each *Verticillium* species.
Table S2. Descriptions of the enriched Pfam domains.
Table S3. Numbers of species-specific genes and numbers of these genes that have homologues in other fungal species.
Data S1
Data S2
Data S3