RESEARCH LETTER – Biotechnology & Synthetic Biology

In silico prediction and characterisation of secondary metabolite clusters in the plant pathogenic fungus *Verticillium dahliae*

Xiaoqian Shi-Kunne¹,§, Roger de Pedro Jové¹,§, Jasper R.L. Depottie¹,²,¹,§, Malaika K. Ebert¹,§, Michael F. Seidl¹,¹,†,¶ and Bart P.H.J. Thomma¹,*,#

¹Laboratory of Phytopathology, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands and ²Department of Crops and Agronomy, National Institute of Agricultural Botany, Huntingdon Road, CB3 0LE Cambridge, United Kingdom

*Corresponding author:* Laboratory of Phytopathology, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. Tel: 0031-317-484536; Fax: 0031-317-483412; E-mail: bart.thomma@wur.nl

§These authors contributed equally to this work.
†Present address: Botanical Institute, University of Cologne, BioCenter, Zülpicher Str. 47a, 50674, Cologne, Germany.
‡Present address: Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands.

One sentence summary: In total, 25 potential SMCs were identified in the genome of the plant pathogenic fungus *Verticillium dahliae*, including loci that can be implicated in DHN-melanin, ferricrocin, triacetyl fusarine and fujikurin production.

Editor: Stefan Olsson

Keywords: Verticillium; pathogen; natural product; genomics; fungi
INTRODUCTION

Filamentous fungi are known for their ability to produce a vast array of distinct chemical compounds that are also known as secondary metabolites (SMs) (Keller, Turner and Bennett 2005). In contrast to primary metabolites, SMs are often considered as non-essential for fungal growth, development or reproduction. However, SMs can be crucial for long-term survival in competitive fungal niches (Fox and Howlett 2008; Ponts 2015; Derntl et al. 2017). SMs produced by plant pathogenic fungi are of particular interest as they may contribute to virulence, leading to crop losses and threatening food security (Ponts 2015; Fusetahelyi, Holb and Pócsi 2015). For example, T-toxin from Cochliobolus heterostrophus, a maize pathogen that caused the worst epidemic in U.S. agricultural history, has been reported to be a crucial pathogenicity factor (Inderbitzin, Asvarak and Turgeon 2010).

Fungal SMs are classified into four main groups based on core enzymes and precursors involved in their biosynthesis: polyketides, non-ribosomal peptides, terpenes and indole alkaloids (Keller, Turner and Bennett 2005). Production of the chemical scaffold of each class requires core enzymes named polyketides synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), terpene cyclases and dimethylallyl tryptophan synthases, respectively. Additionally, hybrid enzymes such as PKS-NRPSs have been identified as builders of structurally complex molecules with combined properties (Boettger and Herwec 2013). PKSs and NRPSs are the most abundant and are extensively studied in fungi (Cox 2007). PKSs can be further divided into three different types (I, II and III), of which type I PKSs and type III PKSs are found in fungi. Type I PKSs are predominant in fungi whereas type III PKSs are found only rarely (Cox 2007; Gallo, Ferrara and Perrone 2013; Hashimoto, Nonaka and Fujii 2014). Genes involved in the synthesis of SMs are frequently located in close proximity to each other, forming so-called secondary metabolism clusters (SMCs) (Keller and Hohn 1997; Brakhage and Schroeckh 2011; Wiemann and Keller 2014). Most of these SMCs contain one biosynthetic core gene that is flanked by transporter proteins, transcription factors and genes encoding tailoring enzymes that modify the SM structure (Keller and Hohn 1997; Keller, Turner and Bennett 2005).

The genomics era has provided new tools to study fungal SMs and their biosynthesis at the whole genome scale (Wiemann and Fischbach 2015). The distinctive traits of gene clusters (e.g. gene distance) and the conserved signatures of core genes (e.g. conserved domains) can be exploited to identify putative loci involved in SM production. Moreover, phylogenetic and comparative genomics analyses are very informative as the number of fungal genomes and characterised SM pathways increases. These two approaches are very helpful to identify gene clusters that are involved in the production of SMs that have been characterised in other fungal species and allow subsequent predictions of identical or related compounds that a particular fungal species might produce (Medema, Takano and Breitling 2013; Cairns and Meyer 2017).

The fungal genus Verticillium contains nine haploid species plus the allodiploid Verticillium longisporum (Inderbitzin et al. 2011; Depoter et al. 2017). These ten species are phylogenetically subdivided into two clades; Flavoxudans and Flavnonexudans (Inderbitzin et al. 2011; Shi-Kunne et al. 2018). The Flavnonexudans clade comprises Verticillium nubilum, Verticillium alfaiae, Verticillium isacii, Verticillium tricus, Verticillium klebahnii and Verticillium zaregamsianum (Inderbitzin et al. 2011). Among these Verticillium spp., V. dahliae is the most notorious plant pathogen that is able to cause disease in hundreds of plant species (Fradin and Thomma 2006; Inderbitzin and Subbarao 2014). Furthermore, V. albo-atrum, V. alfaiae, V. nonalfafae and V. longisporum are pathogenic, albeit with narrower host ranges (Inderbitzin and Subbarao 2014). Although the remaining species V. tricus, V. zaregamsianum, V. nubilum, V. isacii and V. klebahnii have incidentally been reported as plant pathogens, they are mostly considered as saprophytes that thrive on dead organic material and their incidental infections should be seen as opportunistic (Ebihara et al. 2003; Inderbitzin et al. 2011; Gurung et al. 2015). Previously, studies of three genes that are involved in SM biosynthesis in V. dahliae suggested that SMs may play a role in V. dahliae virulence. The deletion mutants of the putative secondary metabolism regulators VdSge1 (Santhanam and Thomma 2012) and VdMcm1 (Xiong et al. 2016) displayed reduced virulence when compared with the wild type V. dahliae strain. Likewise, a reduction in virulence was observed for deletion mutants of the cytochrome P450 monoxygenase VdCPY1 (Zhang et al. 2016), a common tailoring enzyme in SMC production. In this study, we conducted an in silico analysis to unravel the potential secondary metabolism of V. dahliae by making use of the gapless genome assembly of strain JR2 (Fairo et al. 2015).

MATERIALS AND METHODS

Secondary metabolite cluster prediction, annotation and conservation

Putative SMCs were identified with antiSMASH fungal version 4.0.2 (Weber et al. 2015). The predicted borders from antiSMASH were directly used to retrieve all protein sequences contained within the clusters. The Bedtools intersect command (Quinlan and Hall 2010) was used to obtain the file containing the gene locations, followed by gffread from the Cufflinks package (Trapnell et al. 2010) to retrieve the protein sequences. Sub-telomeric regions were defined as 300 kb of the chromosomal ends, as similarly used for other filamentous fungi (McDonagh et al. 2008; Cairns and Meyer 2017). Genes within the genomic range were counted using BioMart from Ensembl (Kersey et al. 2016). A \( \chi^2 \) test was performed to determine the significance of enrichment.

The conservation of predicted SMCs among Verticillium spp. was assessed based on core enzyme conservation, using BLAST + tool protein blast (blastp) (Camacho et al. 2009) on predicted protein databases (e-value < 1 \( \times 10^{-5} \), query coverage > 60% and identity > 50%) (Sbaraini et al. 2017).

Phylogenetic analysis

The previously described type I PKSs, NRPSs and PKS-NRPSs enzymes used in this study for phylogenetic analysis were derived from the curated database of UniProt, SwissProt and literatures (Gallo, Ferrara and Perrone 2013; Yu et al. 2015). The amino acid alignment was built using MAFFT version 7.205 (Katoh and Standley 2013). We used the G-INS-i strategy, global alignment (–globalpair) and 1000 cycles of iterative refinement (–maxiterate 1000). Aligned sequences were visualised with Aliview version 1.20 (Larsson, 2014) and manually curated by removing non-aligned sequences. Preceding the phylogenetic analysis, the alignments were trimmed to...
remove poorly aligned regions using TrimAl version 1.4 (Capella-Gutiérrez, Silla-Martínez and Gabaldón 2009). First, all positions in the alignment with gaps in 90% or more of the sequence were removed (-gt 0.1), followed by the automated1 parameter (-automated1). RaxML version 8.1.1 (Stamatakis 2014) was used to construct Maximum-likelihood phylogenetic tree (-f a). The automated protein model selection (-m PROTGAMMAAUTO) was used applying 100 rapid bootstrapping (-#100). The number of seeds for parsimony inferences and rapid bootstrap analysis was set to 12 345 (-p 12 345 -x12 345, respectively). The output tree (RaxML_bipartitionBranchLabels) was visualised using iTOL webtool version 3.0 (Letunic and Bork 2016).

**Comparative cluster analysis**

Protein sequences of described clusters were blasted (BLASTp, E-value cutoff 1e-5, query coverage > 60% and identity > 25%) against the *V. dahliae* strain JR2 protein database. We considered a cluster to be conserved in *V. dahliae* when at least 50% of the queried proteins from previously described clusters were found in *V. dahliae*.

**Gene expression analysis**

To obtain RNA-seq data for *V. dahliae* grown in culture medium, strain JR2 was grown for 3 days in potato dextrose broth (PDB) in three biological replicates. To obtain RNA-seq data from *V. dahliae* grown in planta, three-year-old Arabidopsis thaliana (Col-0) plants were inoculated with strain JR2. After root inoculation, plants were grown in individual pots in a greenhouse under a cycle of 16 h of light and 8 h of darkness, with temperatures maintained between 20 and 22°C during the day and a minimum of 15°C overnight. Three pooled samples (10 plants per sample) of complete flowering stems were used for total RNA extraction. Total RNA was extracted based on TRizol RNA extraction (Simms, Cizziel and Chomczynski 1993), cDNA synthesis, library preparation (TruSeq RNA-Seq short insert library) and Illumina sequencing (single-end 50 bp) was performed at the Beijing Genome Institute (BGI, Hong Kong, China). In total, ~2 Gb and ~1.5 Gb of filtered reads were obtained for the *V. dahliae* samples grown in culture medium and in planta, respectively. RNAseq data were submitted to the SRA database and can be accessed through BioProject accession PRJNA473305.

The RNA sequencing reads were mapped to their previously assembled genomes using the Raubread package (v1.32.2) in R (Liao et al. 2013). The comparative transcriptomic analysis was performed with the package edgeR in R (v3.4.3) (Robinson, McCarthy and Smyth 2010; McCarthy, Chen and Smyth 2012). Genes are considered differentially expressed when *P*-value < 0.05 with a log2-fold-change ≥ 1. *P*-values were corrected for multiple comparisons according to Benjamini and Hochberg (1995).

**RESULTS AND DISCUSSION**

The *V. dahliae* strain JR2 genome contains 25 putative secondary metabolite clusters

To assess the potential secondary metabolism of *V. dahliae* strain JR2, we mined its genome sequence to predict SMCs using antiSMASH (Weber et al. 2015). A total of 25 putative SMCs were predicted, containing a total of 364 genes within their boundaries (Fig. 1 and Table 1). The putative SMCs were classified as nine type I PKs, one type III PKs, one PKS-NRPS, three NRPSs and four terpenes. Seven clusters were classified as ‘other’, a generic class of SMCs containing core enzymes with unusual domain architecture, also known as non-canonical. We found that each of these SMCs contains one core gene. Disrupted SMCs frequently occur in the genomes of filamentous fungi (Collemare et al. 2014). However, none of the clusters identified in *V. dahliae* strain JR2 is evidently disrupted, as all identified clusters have no insertions of transposable elements and include genes encoding tailoring enzymes such as methyltransferases, cytochrome P450 or dehydrogenases (Cacho, Tang and Chooi 2015). Several clusters also comprise transporter and transcription factor encoding genes that might be involved in SM secretion and local gene cluster regulation respectively (Collemare et al. 2014) (Table 1). Collectively, these results suggest that all the analysed SMCs of *V. dahliae* strain JR2 are potentially functional.

In several species, SM genes are enriched at chromosomal ends adjacent to the telomeres (Cairns and Meyer 2017; Farmer 2007; McDonagh et al. 2008). Thus, we assessed whether the SMCs of *V. dahliae* are located in sub-telomeric regions, here defined as within 300 kb from the chromosomal end. We found that 36% of the predicted clusters are in sub-telomeric regions (Fig. 1A and Table 1). As these sub-telomeric regions harbour 13.8% of the total gene repertoire (1606 genes in total) (Table 1), they are significantly enriched ($\chi^2$-test, $P < 0.0001) in secondary metabolism genes (Fig. 1A).

To check whether the SMCs identified in *V. dahliae* strain JR2 are also present in other *V. dahliae* strains, we assessed the presence/absence of the core SM enzymes in 22 *V. dahliae* strains (Klosterman et al. 2011; Faino et al. 2015; Kombrink et al. 2017; Depotter et al. 2018). Among these 22 strains is the gapless genome assembly of strain VdLs17 (Klosterman et al. 2011; Faino et al. 2015) and the nearly complete genome assemblies of strains CQ2 and 85S (Depotter et al. 2018). The remaining genome assemblies are considerably fragmented, with over 500 contigs for each of the assemblies. Nevertheless, we found that all but one of the core enzymes is present in all 22 strains, except for VdPks8 which was next to JR2 only found in VdLs17, CQ2 and 85S. However, the absence of VdPks8 from the other strains may be due to the fragmented genome assemblies. Subsequently, we assessed the genome assemblies of strains VdLs17, CQ2 and 85S for the presence of complete clusters, revealing that all SMCs identified in *V. dahliae* strain JR2 are also found in the genomes of these three strains, therefore suggesting that SMCs are highly conserved in *V. dahliae* strains.

To examine whether the SMCs identified in *V. dahliae* strain JR2 are also present in other Verticillium spp., we queried core enzymes of each cluster using BLAST against the proteomes of previously published Verticillium spp. (Depotter et al. 2017; Shi-Kunne et al. 2018). In total, 12 SMC core enzymes are present in all Verticillium spp., nine of which showed considerable sequence conservation (~80% sequence identity) (Fig. 1B). The other 15 core enzymes are not present in all species, but display a mosaic presence/absence pattern with presence in at least two other species. Of these, VdPks7 is conserved in all species except for *V. longisporum* (Fig. 1B). VdPks5 is only conserved in the closely related species *V. alfalfae* and *V. nonalfalfae* and VdPks5 is conserved in all pathogenic species except *V. albo-atrum*. Interestingly, the VdPks8 core enzyme was only found in a single copy in the hybrid *V. longisporum* genome, presumably derived from its *V. dahliae* progenitor. Thus, based on the widespread presence of these core genes within the Verticillium genus, we predict that most of the SMCs are conserved throughout this genus.
**Phylogenomic analysis of V. dahliae secondary metabolite core enzymes**

In ascomycetes, type I PKSs, NRPSs and PKS-NRPSs are known to produce most of the SMs that are involved in virulence (Pusztahelyi, Holb and Pócsi 2015). Thus, we focused on identifying putative functions of type I PKSs, NRPSs and PKS-NRPSs in V. dahliae with a phylogenomics approach. To this end, we aligned KS domains of V. dahliae PKSs to KS domains of functionally described PKSs, and subsequently constructed a phylogenetic tree that comprises three major clades that correspond to the NR-PKSs, PR-PKSs and HR-PKSs, respectively (Fig. 2). The NR-PKS clade contains two predicted V. dahliae PKSs, VdPks2 and VdPks3. VdPks2 clusters with PKSs that have been implicated in dihydroxynaphthalene (DHN)-melanin formation (Tsuji et al. 2002; Yu et al. 2015). VdPks3 clustered with the Orsellinic acid synthase
Table 1. Predicted secondary metabolite gene clusters (SMCs) in V. dahliae strain JR2

| SM class | Predicted key biosynthetic gene | Gene name | Sub-termineric | Transporter | Transcription factor | Cluster location | No. of genes |
|----------|---------------------------------|-----------|---------------|-------------|----------------------|-----------------|-------------|
| PKS-NRPS | VDAGJR2Chr1g1100                | VdPks1    | NO            | YES         | YES                  | 1.3 545 717:3 591 771 | 16          |
|          | VDAGJR2Chr1g1588                | VdPks2    | NO            | NO          | YES                  | 1.5 097 044:5 143 749 | 15          |
|          | VDAGJR2Chr2g0045                | VdPks3    | YES           | YES         | YES                  | 2.92 065:145 583 18 | 18          |
|          | VDAGJR2Chr2g0695                | VdPks4    | NO            | YES         | YES                  | 2.2 119 540:2 167 273 | 14          |
|          | VDAGJR2Chr3g0719                | VdPks5    | NO            | YES         | YES                  | 2.2 182 651:2 230 558 | 20          |
|          | VDAGJR2Chr3g0903                | VdPks6    | NO            | YES         | YES                  | 3.316 303:358 028 14 | 14          |
|          | VDAGJR2Chr4g1125                | VdPks7    | NO            | NO          | NO                   | 4.3 709 413:6 756 014 | 17          |
|          | VDAGJR2Chr5g1015                | VdPks8    | YES           | YES         | YES                  | 5.384 693:3 432 789 | 11          |
|          | VDAGJR2Chr5g1031                | VdPks9    | YES           | YES         | YES                  | 8.3 058 903:3 104 627 | 21          |
| T3PKS4   | VDAGJR2Chr1g23880               | VdHyb1    | NO            | YES         | YES                  | 1.7 618 782:7 670 067 | 20          |
|          | VDAGJR2Chr3g1324                | VdNprs1   | YES           | YES         | YES                  | 3.4 066 099:4 121 170 | 19          |
|          | VDAGJR2Chr6g0600                | VdNprs2   | NO            | YES         | YES                  | 6.1 768 180:1 819 114 | 17          |
|          | VDAGJR2Chr7g1025                | VdNprs3   | YES           | YES         | YES                  | 7.3 179 393:3 227 591 | 13          |
|          | VDAGJR2Chr4g0955                | VdT3pks   | YES           | YES         | YES                  | 4.3 205 819:3 247 084 | 14          |
|         | VDAGJR2Chr1g1245                | VdTerp1   | NO            | NO          | NO                   | 1.4 010 610:4 031 875 | 7           |
|         | VDAGJR2Chr1g1723                | VdTerp2   | NO            | NO          | NO                   | 1.5 472 865:5 493 695 | 8           |
|         | VDAGJR2Chr2g0913                | VdTerp3   | NO            | NO          | NO                   | 2.2 797 712:2 819 868 | 10          |
|         | VDAGJR2Chr7g0295                | VdTerp4   | NO            | NO          | YES                  | 7.844 777:866 572 5 | 7           |
| Other    | VDAGJR2Chr2g0388                | Other1    | NO            | YES         | YES                  | 2.1 232 486:1 276 127 | 15          |
|          | VDAGJR2Chr4g0468                | Other2    | NO            | YES         | YES                  | 4.1 587 321:1 632 662 | 16          |
|          | VDAGJR2Chr5g1148                | Other3    | YES           | YES         | YES                  | 5.3 902 983:3 948 100 | 14          |
|          | VDAGJR2Chr5g1176                | Other4    | YES           | YES         | YES                  | 5.3 993 259:4 036 240 | 15          |
|          | VDAGJR2Chr6g0066                | Other5    | YES           | YES         | YES                  | 6.150 534:194 469 14 | 14          |
|          | VDAGJR2Chr6g0109                | Other6    | YES           | YES         | YES                  | 6.289 905:335 580 18 | 18          |
|          | VDAGJR2Chr8g0117                | Other7    | YES           | YES         | YES                  | 8.289 821:333 602 13 | 13          |

1 T1PKS = type 1 polyketide synthase.
2 PKS-NRPS = hybrid polyketide synthase-non-ribosomal peptide synthase.
3 NRPS = non-ribosomal peptide synthase.
4 T3PKS = type 3 polyketide synthase.

Sub-terminalic clusters were defined as any cluster predicted to reside within 300 kb of a chromosome end.

OpS1, which is involved in production of the toxic metabolite oosporein by the entomopathogenic fungus Beauveria bassiana (Feng et al. 2015). The HR-PKS clade showed that four out of the seven predicted V. dahliae HR-PKSs (VdPks1, VdPks7, VdPks8 and VdPks9) grouped with previously described enzymes. VdPks7 and VdPks8 clustered with the fumagillin synthase from Aspergillus fumigatus and fujikurin synthase from Fusarium fujikuroi, respectively (Lin et al. 2013; von Bargen et al. 2015; Niehaus et al. 2016). VdPks1 and VdPks9 clustered with the T-toxin synthase PKS2 from C. heterostrophus (Inderbitzin, Asvarak and Turgeon 2010). VdPks4 grouped in a clade that contains sdnO and PKS1 synthase, which are involved in the production of the sor dinar by Sordaria araneosa (Kudo et al. 2016), an antifungal agent that inhibits protein synthesis in fungi by stabilising the ribosome/EF2 complex (Justice et al. 1998), and in T-toxin production in C. heterostrophus (Inderbitzin, Asvarak and Turgeon 2010), respectively. VdPks5 is in a clade that only contains FUB1 fusaric acid synthase orthologs of three Fusarium spp. (Brown et al. 2007; Sieber et al. 2014; Oide et al. 2015). VdNprs2 clusters with VdPks4, which is responsible for the synthesis of the extracellular siderophore triacylfusarinine C (TAF) from A. fumigatus (Schreftl et al. 2007). The clade that contains VdNprs3 has low bootstrap values and long branches, indicating considerable divergence of this enzyme (Fig. 3). Thus, only two of the V. dahliae NRPS core enzymes (VdNprs1 and VdNprs2) group with previously characterised enzymes.

The fusion of PKS and NRPS domains results in PKS-NRPS enzymes that end out of their structural complexity (Boettger and Hertweck 2013). In the genome of V. dahliae (strain JR2), only one PKS-NRPS (VdHyb1) was detected. Since it contains a KS domain characteristic for HR-PKSs and an A-domain as commonly observed in NRPSs, we included VdHyb1 in both phylogenetic trees. In the PKS phylogeny (Fig. 2) VdHyb1 was found in the same clade (38%, bootstrap value) as the PKS-NRPSs chaetoglobosin A synthase from Chaetomium globosum and the avirulence protein ACE1 from Magnaporthe grisea (Collemare et al. 2008; Ishiuchi et al. 2013). In contrast, VdHyb1 did not cluster with any previously characterised PKS-NRPSs in the A-domain phylogenetic tree (Fig. 3).
Figure 2. Phylogenetic tree of type I PKS and PKS-NRPS enzymes. KS domains of PKS and PKS-NRPS enzymes were aligned to construct the maximum likelihood tree with 100 bootstrap replicates. The chicken fatty acid synthase (Chicken FAS) sequence was used as outgroup. Only bootstrap values above 30 are shown below the branches. *V. dahliae* KS domains are highlighted in blue and indicated with an arrow. Protein codes correspond to Uniprot IDs.
Figure 3. Phylogenetic tree of NRPS and PKS-NRPS enzymes. A-domains of NRPS and PKS-NRPS enzymes were aligned to construct the maximum likelihood tree with 100 bootstrap replicates. Only bootstrap values over 30 are shown below the branches. V. dahliae A-domains are highlighted in blue and indicated with an arrow.
Figure 4. Synteny of conserved SMCs in V. dahliae. V. dahliae putative clusters were compared to previously described clusters. Ensembl gene IDs are shown above or below the genes. (A) DHN-melanin, (B) Fujikurins, (C) Ferricrocin, (D) Triacetyl fusarinine.

Comparative analysis of gene clusters

Based on the phylogenetic analyses, 11 core enzymes (VdPks1, VdPks2, VdPks3, VdPks4, VdPks5, VdPks7, VdPks8, VdPks9, VdNrps1, VdNrps2 and VdHyb1) were identified that cluster with previously characterised enzymes from other fungal species (Figs. 2 and 3). Subsequently, we queried for the other genes besides the core genes from these previously characterised clusters in other fungal species to find homologs in the corresponding V. dahliae clusters. However, only the VdPks2, VdPks8, VdNrps1 and VdNrps2 clusters of V. dahliae share more homologs (more than 50% of the whole cluster) in addition to the core genes with other fungal species. The remaining clusters contain less than 50% of genes that share homologs with other fungal species. In other fungi, conserved gene clusters of VdPks2, VdPks8, VdNrps1 and VdNrps2 are responsible for the biosynthesis of DHN-melanin, fujikurins, ferricrocin and TAFC, respectively (Tsuji et al. 2002; Tobiasen et al. 2007; Sieber et al. 2014; von Bargen et al. 2015; Niehaus et al. 2016).

The gene cluster for DHN-melanin biosynthesis in the fungal pathogen C. lagenarium comprises six genes, including three functionally characterised genes encoding PKS ClPKS1, reductase T4HR1 and transcription factor cmr1. Moreover, two genes (scytalone dehydratase SCD1 and THR1 reductase) residing at another chromosome were identified to be involved in the biosynthesis of DHN-melanin as well (Tsuji et al. 2002) (Fig. 4A). We observed amino acid identities of 70%–90% between ClPKS1, T4HR1, SCD1 and THR1 of C. lagenarium and their orthologs in V. dahliae. The transcription factor CMR1 only shares 55% identity with its counterpart in C. lagenarium.

VdPks8 is an ortholog of the fujikurin synthase gene (FfuPks19) of F. fujikuroi. The fujikurin cluster contains six genes (von Bargen et al. 2015; Niehaus et al. 2016), four of which have homologs in the VdPks8 locus in V. dahliae. The homolog of the MFS transporter FFUJ12 242 of F. fujikuroi was found on a different chromosome in V. dahliae, and the cytochrome P450 gene in F. fujikuroi has no V. dahliae homolog (Fig. 4B). Interestingly, the VdPks8 locus contains two other genes that are annotated as cytochrome P450 and MFS transporter, but these genes were not detected as homologs of the genes in the FfuPks19 cluster (Fig. 4B).

The biosynthesis of ferricrocin requires two genes that are located at the same locus in the genome of F. pseudograminearum, which encode an L-ornithine N5-oxygenase (SIDA) and an NRPS (FpNRPS2) (Tobiasen et al. 2007; Sieber et al. 2014). Similarly, in V. dahliae homologs of L-ornithine N5-oxygenase (SIDA) and NRPS (FpNRPS2) genes are located next to each other. Moreover, homologs of a proteasome subunit, a transcription factor and two uncharacterised genes in the same cluster of F. pseudograminearum were found in V. dahliae. In addition, genes encoding an MFS transporter and an oxireductase in F. pseudograminearum have no homologs in V. dahliae (Fig. 4B).

VdNrps2 is an ortholog of the extracellular siderophore TFAC synthase gene NRPS4 in A. fumigatus, which belongs to a cluster of two genes (NRPS4 and sidF) (Schrettl et al. 2007). Another two extracellular siderophore TFAC synthase genes (sidG and sidA) are located at another chromosome (Schrettl et al. 2007). Except for sidG, all described genes have homologs in V. dahliae (Fig. 4D).

Gene cluster expression analyses

To identify differentially expressed SMCs, we queried RNA-seq data sets from V. dahliae grown in vitro in PDB, Murashige and Skoog (MS) medium and xylem sap that was harvested from
healthy tomato plants. In order for a gene cluster to qualify as differentially expressed, genes within the clusters must show similar expression patterns and the expression must differ between conditions. However, we found that the majority of gene clusters do not show such expression patterns, meaning that the genes within a cluster do not appear to be co-regulated. Only the clusters VdPks2, VdNrps2, VdNrps3, VdTerp2, and Other4 show co-regulated expression patterns, since the majority of genes within those clusters (>75%) have similar relative expression levels in at least two conditions (Fig. 5A). Furthermore, these five clusters show differential expression (>2 fold) under those two conditions. Interestingly, all five clusters show higher relative expression in xylem sap when compared with PDB and MS medium.

To assess the potential of SM clusters to play a role in pathogenicity, we assessed transcriptome (RNA-seq) data of *V. dahliae* during colonisation of *A. thaliana* plants. We found that the majority of SMC genes are not expressed in planta, although eight core SMC genes are expressed under those conditions (Fig. 5B). We therefore compared the relative expression levels of each core gene in planta and in vitro (PDB), whereas the Other4 core gene was found to be induced in planta, the core genes of VdPks2, VdNrps1, VdNrps2 VdNrps3, VdTerp1, VdTerp3 and Other5 were repressed when compared with the expression in in vitro-cultured mycelium (Fig. 5B). This finding may suggest that Other4 plays a role during colonisation of *A. thaliana*. However, it cannot be excluded that the six core genes that are down-regulated in planta still play a role during host colonisation as well since residual expression remains.

**CONCLUSIONS**

In this study, we have used an *in silico* approach to identify 25 putative SMCs in the genome of *V. dahliae* strain JR2, all of which appear complete and thus potentially functional. Our predictions state that two putative siderophores, ferricrocin and TAFc, DHN-melanin and fujikurin compounds may belong to the active SM repertoire of *V. dahliae*.

**ACKNOWLEDGEMENTS**

Work in the laboratories of B.P.H.J.T. and M.F.S. is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands Organisation of Scientific Research (NWO).

**Conflict of interest.** None declared.
REFERENCES

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

Boettger D, Hertweck C. Molecular diversity sculpted by fungal PKS-NRPS hybrids. ChemBioChem 2013;14:28–42.

Brakhage AA, Schroech V. Fungal secondary metabolite strategies to activate silent gene clusters. Fungal Genet Biol 2011;48:15–22.

Brown DW, Lee S-H, Kim L-H et al. Identification of a 12- gene fusaric acid biosynthetic gene cluster in Fusarium species through comparative and functional genomics. Mol Plant-Microbe Interact 2015;28:319–32.

Cacho RA, Tang Y, Chooi Y-H. Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. Front Microbiol 2015;5:774.

Cairns T, Meyer V. In silico prediction and characterization of secondary metabolite biosynthetic gene clusters in the wheat pathogen Zymoseptoria tritici. BMC Genomes 2017;18:631.

Camacho C, Coulouris G, Avagyan V et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009;25:1972–3.

Collemare J, Griffiths S, Iida Y et al. Secondary metabolism and biotrophic lifestyle in the tomato pathogen Cladosporium fulvum. PLoS One 2014;9:e85877.

Collemare J, Pianetti M, Houle A et al. Magnaporthe grisea avirulence gene ACE1 belongs to an infection-specific gene cluster involved in secondary metabolism. New Phytol 2008;179:196–208.

Cox RJ. Polyketides, proteins and genes in fungi: programmed nanomachines begin to reveal their secrets. Org Biomol Chem 2007;5:2010–26.

Deprotter JRL, Seidl MF, van den Berg GCM et al. A distinct and genetically diverse lineage of the hybrid fungal pathogen Verticillium longisporum population causes stem stripeing in British oilseed rape. Environ Microbiol 2017;19:3997–4009.

Deprotter JRL, Shi-Kunne X, Missonnier H et al. Dynamic virulence-related regions of the fungal plant pathogen Verticillium dahliae display remarkably enhanced sequence conservation. bioRxiv 2018:277558.

Derni C, Kluger B, Bueschl C et al. Transcription factor Xpp1 is a switch between primary and secondary fungal metabolism. Proc Natl Acad Sci U S A 2017;114:E560–9.

Ebihara Y, Uematsu S, Nagao H et al. First report of Verticillium tricorpus isolated from potato tubers in Japan. Mycoscience 2003;44:481–8.

Faino L, Seidl MF, Datema E et al. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. MBio 2015;6:e00936–15.

Farman ML. Telomeres in the rice blast fungus Magnaporthe oryzae: the world of the end as we know it. FEMS Microbiol Lett 2007;273:125–32.

Feng P, Shang Y, Cen K et al. Fungal biosynthesis of the bibenzoquinone oosporein to evade insect immunity. Proc Natl Acad Sci U S A 2015;112:11365–70.

Fox EM, Howlett BJ. Secondary metabolism: regulation and role in fungal biology. Curr Opin Microbiol 2008;11:481–7.

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

Gallo A, Ferrara M, Perrone G. Phylogenetic study of polyketide synthases and nonribosomal peptide synthetases involved in the biosynthesis of mycotoxins. Toxins 2013;5:717–42.

Gurung S, Short DPG, Hu X et al. Host range of Verticillium isacii and Verticillium klebsianii from artichoke, spinach, and lettuce. Plant Dis 2015;99:933–8.

Hashimoto M, Nonaka T, Fuji i I. Fungal type III polyketide synthases. Nat Prod Rep 2014;31:1306–17.

Inderbitzin P, Asvarak T, Turgeon BG. Six new genes required for production of T-toxin, a polyketide determinant of high virulence of Cochliobolus heterostrophus to maize. Mol Plant-Microbe Interact 2010;23:458–72.

Inderbitzin P, Bostock RM, Davis MR et al. Phylogenetics and taxonomy of the fungal vascular wilt pathogen Verticillium, with the descriptions of five new species. PLoS One. 2011;6:e28341.

Inderbitzin P, Subbarao KV. Verticillium systematics and evolution: how confusion impedes Verticillium wilt management and how to resolve it. Phytopathology 2014;104:564–74.

Ishiuchi K, Nakazawa T, Yagishita F et al. Combinatorial generation of complexity by redox enzymes in the chaetoglobosin A biosynthesis. J Am Chem Soc 2013;135:7371–7.

Justice MC, Hsu MJ, Tse B et al. Elongation factor 2 as a novel target for selective inhibition of fungal protein synthesis. J Biol Chem 1998;273:3148–51.

Katoh K, Stanley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2012;30:772–80.

Keller NP, Hohn TM. Metabolic pathway gene clusters in filamentous fungi. Fungal Genet Biol 1997;21:17–29.

Keller NP, Turner G, Bennett JW. Fungal secondary metabolism—from biochemistry to genomics. Nat Rev Microbiol 2005;3:937.

Kersey PJ, Allen JE, Armean I et al. Ensembl genomes 2016: more genomes, more complexity. Nucleic Acids Res 2016;44:D574–80.

Klosterman SJ, Subbarao KV, Kang S et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLOS Pathog 2011;7:e1002137.

Kombrink A, Rovenich H, Shi-kunne X et al. Verticillium dahliae lysM effectors differentially contribute to virulence on plant hosts. Mol Plant Pathol 2017;8:596–608.

Kudo F, Matsuura Y, Hayashi T et al. Genome mining of the sor dinarin biosynthetic gene cluster from Sordaria araneosa Cain ATCC 36386: characterization of cycloaraneoeseene synthase and GDP-6-deoxyaltrose transferase. J Antibi 2016;69:541.

Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 2014;30:3276–8.

Letunic I, Bork P. Interactive tree of life (iTOP) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 2016;44:W242–5.

Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 2013;41:e108-

Lin H-C, Chooi Y-H, Dhingra S et al. The fumagillin biosynthetic gene cluster in Aspergillus fumigatus encodes a cryptic terpene cyclase involved in the formation of β-transgamotone. J Am Chem Soc 2013;135:4616–9.

McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 2012;40:4288–97.

McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics 2009;25:765–71.

McDonagh A, Fedorova ND, Crabtree J et al. Sub-telomere directed gene expression during initiation of invasive aspergillosis. PLOS Pathog 2008;4:e1000154.
Medema MH, Fischbach MA. Computational approaches to natural product discovery. Nat Chem Biol 2015;11:639.

Medema MH, Takano E, Breitling R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol Biol Evol 2013;30:1218–23.

Niehaus E-M, Münsterkötter M, Proctor RH et al. Comparative “Omics” of the Fusarium fujikuroi species complex highlights differences in genetic potential and metabolite synthesis. Genome Biol Evol 2016;8:3574–99.

Oide S, Berthiller F, Wiesenberger G et al. Individual and combined roles of malonichrome, ferricrocin, and TAF siderophores in Fusarium graminearum pathogenic and sexual development. Front Microbiol 2015;5:759.

Ponto N. Mycotoxins are a component of Fusarium graminearum stress-response system. Front Microbiol 2015;6:1234.

Pusztahelyi T, Holb I, Pócsi I. Secondary metabolites in fungus-plant interactions. Front Plant Sci 2015;6:573.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2.

Robinson MD, McCarthy Dj, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40.

Santhanam P, Thomma BPHJ. Verticillium dahliae Sge1 differentially regulates expression of candidate effector genes. Mol Plant-Microbe Interact 2012;26:249–56.

Sbaraini N, Andreis FC, Thompson CE et al. Genome-wide analysis of secondary metabolite gene clusters in Ophiostoma ulmi and Ophiostoma novo-ulmi reveals a fujikurin-like gene cluster with a putative role in infection. Front Microbiol 2017;8:1063.

Schrettl M, Bignell E, Kragl C et al. Distinct roles for intra- and extracellular siderophores during Aspergillus fumigatus infection. PLOS Pathog 2007;3:e128.

Shi-Kunne X, Faino L, van den Berg GCM et al. Evolution within the fungal genus Verticillium is characterized by chromosomal rearrangement and gene loss. Environ Microbiol 2018;20:1362–73.

Sieber CMK, Lee W, Wong P et al. The Fusarium graminearum genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. PLoS One 2014;9:e110311.

Simms D, Cizziel PE, Chomczynski P. TRizol: a new reagent for optimal single-step isolation of RNA. Focus 1993;15:532–5.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.

Tobiasen C, Aahman J, Ravnholt KS et al. Nonribosomal peptide synthetase (NPS) genes in Fusarium graminearum, F. culmorum and F. pseudograminearum and identification of NPS2 as the producer of ferricrocin. Curr Genet 2007;51:43–58.

Trapnell C, Williams BA, Pertea G et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511.

Tsuij G, Kenmochi Y, Takano Y et al. Novel fungal transcriptional activators, Cmr1p of Colletotrichum lagenarium and Pig1p of Magnaporthe grisea, contain Cys2His2 zinc finger and Zn(II)Cys6 binuclear cluster DNA-binding motifs and regulate transcription of melanin biosynthesis. Mol Microbiol 2002;38:940–54.

von Bargen KW, Niehaus E-M, Krug I et al. Isolation and structure elucidation of fujikurins A–D: products of the PKS19 gene cluster in Fusarium fujikuroi. J Nat Prod 2015;78:1809–15.

Weber T, Blin K, Duddela S et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 2015;43:W237–43.

Wiemann P, Keller NP. Strategies for mining fungal natural products. J Ind Microbiol Biotechnol 2014;41:301–13.

Xiong D, Wang Y, Tian L et al. MADS-Box transcription factor VdMcm1 regulates conidiation, microsclerotia formation, pathogenicity, and secondary metabolism of Verticillium dahliae. Front Microbiol 2016;7:1192.

Yu X, Huo L, Liu H et al. Melanin is required for the formation of the multi-cellular conidia in the endophytic fungus Pestalotiopsis microspora. Microbiol Res 2015;179:1–11.

Zhang DD, Wang XY, Chen JY et al. Identification and characterization of a pathogenicity-related gene VdCYP1 from Verticillium dahliae. Sci Rep 2016;6:27979.