RESEARCH LETTER

Structural and phylogenetic analysis of the rDNA intergenic spacer region of *Verticillium dahliae*

Ioannis A. Papaioannou, Chrysoula D. Dimopoulou & Milton A. Typas

Department of Genetics & Biotechnology, Faculty of Biology, National and Kapodistrian University of Athens, Athens, Greece

Abstract

The nuclear ribosomal intergenic spacer (IGS) region was structurally analyzed and exploited for molecular discrimination and phylogenetic analysis of vegetative compatibility groups (VCGs) of *Verticillium dahliae*. A structural study of 201 available IGS sequences of the fungus was performed, and four classes of ubiquitous repetitive elements, organized in higher-order repetitive structures or composite blocks, were detected in a variable IGS subregion. This subregion was amplified from an international collection of 59 *V. dahliae* isolates covering all VCGs, together with nine representative *V. albo-atrum* and *V. longisporum* isolates, and sequenced. Structural and phylogenetic analyses of the sequences of this polymorphic IGS subregion were consistently informative and allowed the identification of two main lineages in *V. dahliae*, that is, clade I including VCGs 1A, 1B, 2A, 4B, and 3 and clade II containing VCGs 2B, 4A, and 6. Analysis of IGS sequences proved a highly suitable molecular tool for (a) rapid interspecific differentiation, (b) intraspecific discrimination among VCGs of *V. dahliae*, facilitating high-throughput VCG confirmation and prediction/profiling, and (c) phylogenetic analysis within and among *V. dahliae* VCGs.

Introduction

Verticillium dahliae is a cosmopolitan soil-borne ascomycete causing wilt diseases on more than 400 agronomically important plant species. A thorough understanding of its population biology is predicted to facilitate disease management (Pegg & Brady, 2002; Atallah *et al.*, 2010). Numerous studies have classified isolates of *V. dahliae* into vegetative compatibility groups (VCGs) 1, 2, 3, 4, and 6 (of these, VCGs 1, 2, and 4 have been further subdivided depending on the vigor of heterokaryotic reactions). Because no sexual stage has ever been observed for *Verticillium* species, these fungi are predicted to exchange genetic material only through heterokaryon formation and the parasexual cycle, and thus, different VCGs can be assumed to represent genetically isolated intraspecific groups (Rowe, 1995; Katan, 2000; Bhat *et al.*, 2003; Jiménez-Díaz *et al.*, 2006). Although VCG classification is a convenient tool for the characterization of diversity in fungal populations, it is insufficient for the determination of genetic relatedness between isolates of *V. dahliae*, due to inherent methodological limitations, the small number of VCGs

defined – hence low resolution for *V. dahliae* populations – and the unknown nature of the genetic system governing vegetative incompatibility in this fungus (Joaquim & Rowe, 1990; Leslie, 1993; Daayf *et al.*, 1995; Rowe, 1995). In recent years, several molecular techniques including amplified fragment length polymorphism (AFLP) analyses (Collado-Romero *et al.*, 2006), PCR-based molecular marker tools (Collado-Romero *et al.*, 2009; Papaioannou *et al.*, 2013), and multilocus phylogenetic studies (Collado-Romero *et al.*, 2008; Martin, 2010) have been employed to study the genetic relationships within and among VCGs of *V. dahliae*. Notably, individual and even multilocus genealogies provided little resolution at the VCG level, due to the very low level of polymorphism observed in the DNA regions that were investigated (Collado-Romero *et al.*, 2008). Thus, sequences of novel DNA regions that provide more phylogenetic information and better discriminatory resolution for VCGs are needed for the reliable high-throughput characterization of *V. dahliae* populations at the VCG level.

The nuclear ribosomal RNA gene complex (rDNA) has been the most popular genetic region used for fungal

MICROBIOLOGY LETTERS

FEMS

species discrimination and phylogenetics. However, no significant variability has been detected within the small and large ribosomal subunit RNA genes in *V. dahliae* populations, and the internal transcribed spacer (ITS) region is remarkably conserved among VCGs of the fungus (Pramateftaki *et al.*, 2000; Collado-Romero *et al.*, 2008; Papaioannou *et al.*, 2013). In contrast, the intergenic spacer (IGS) region was shown to be highly polymorphic even at the intraspecific level for several fungi, including *V. dahliae*, thus providing a useful tool for taxonomy and phylogeny (Morton *et al.*, 1995; Pramateftaki *et al.*, 2000; Pantou *et al.*, 2003; Hong *et al.*, 2005; Mbofung *et al.*, 2007; Tran *et al.*, 2013).

The purpose of the present study was to assess whether IGS sequences can be used for high-resolution molecular discrimination and phylogenetic analyses among VCGs of *V. dahliae*. To this end, a structural study of publicly available IGS sequences from *V. dahliae* isolates was performed, and the most variable IGS subregion was determined. This subregion was amplified from a collection of diverse *V. dahliae* isolates covering all VCGs, sequenced, and structurally and phylogenetically analyzed. We demonstrate here that IGS analysis is a highly suitable molecular tool for VCG prediction and profiling of *V. dahliae* populations, as well as for phylogenetic studies of the fungus.

Materials and methods

Fungal isolates and culture conditions

The IGS sequences of 201 *V. dahliae* isolates from different hosts and geographic origins (Table S1), of which the VCGs of only 11 isolates were known, were retrieved from GenBank (National Center for Biotechnology Information, USA; website http://www.ncbi.nlm.nih.gov/genbank/) and used for structural analysis. The IGS sequences of 59 *V. dahliae* (from diverse VCGs, host sources, and origins), four *V. albo-atrum*, and five *V. longisporum* isolates were employed in this work for VCG-related structural and phylogenetic analyses (Table 1). The maintenance of monoconidial cultures of all isolates and the culture conditions have been described previously (Papaioannou *et al.*, 2013).

DNA extraction, PCR, cloning, and sequencing

Total DNA of all isolates was extracted according to Typas *et al.* (1992) and stored at −20 °C until further use. The IGS rDNA of each isolate was amplified with PCR using primers CNL12 and CNS1 (White *et al.*, 1990) and Kapa HiFi DNA polymerase (Kapa Biosystems, Woburn, MA), according to the manufacturer's suggestions and

procedures described previously (Papaioannou *et al.*, 2013). Because direct sequencing of PCR amplicons occasionally produced sequences of unsatisfactory quality during preliminary experiments (probably due to the *in vitro* formation of secondary structures or the presence of multiple IGS haplotypes within individuals), a cloning strategy was adopted. In detail, all PCR products were recovered from agarose gels, purified with the NucleoSpin Gel and PCR Clean-Up kit (Macherey–Nagel, Düren, Germany), and cloned using the CloneJET PCR cloning kit (Fermentas, Thermo Fisher Scientific, Waltham, MA). For each isolate, five bacterial clones were picked up after transformation and checked for insert size with colony PCR using universal vector primers, according to the manufacturer's suggestions. A single clone exhibiting the anticipated insert size was selected for bidirectional sequencing. Plasmids were purified using the NucleoSpin Plasmid kit (Macherey–Nagel) before used as templates for sequencing reactions. Automated sequencing (based on the Sanger method) was carried out according to Papaioannou *et al.* (2013) with primers VdIGSpF2 (5′-TCAATTCCCGGG TAGGTGGTCTCT-3′) and VdIGSpR (5′-ACGCCGCT GCAGCCGAAAGT-3′), which anneal to conserved sites of the IGS region. All sequences were deposited at the GenBank database (NCBI), with accession numbers KC152269-KC152319.

Structural and phylogenetic analyses

Deposited DNA sequences were retrieved from GenBank (NCBI), and DNA similarity searches were performed with Basic Local Alignment Search Tool (BLAST 2.2.27+; Altschul *et al.*, 1997). Multiple alignments of nucleotide sequences were prepared using the MUSCLE algorithm as implemented in MEGA 5.05 (Tamura *et al.*, 2011), with minor manual editing after visual inspection. The program GENEQUEST of the software package Lasergene 6 (DNAstar, Madison, WI) was used for structural analysis of the IGS rDNA region. Phylogenetic analysis was performed using the maximum-likelihood (ML) method in software MEGA 5.05 (Tamura *et al.*, 2011) and the Bayesian inference (BI) method in program MRBAYES 3.2.1 (Ronquist *et al.*, 2012). The Bayesian information criterion and the corrected Akaike information criterion scores were evaluated for selection of the substitution model best fitting the dataset for ML analysis (Tamura's three-parameter model with invariant sites) in MEGA 5.05. Reliability of nodes of the topologies was assessed by bootstrap analysis based on 1000 replications. For BI analysis, random starting trees were used, and the burn-in period was set to 500 000 generations as this was found to be clearly sufficient for the likelihood and the model parameters to reach equilibrium. After the burn-in

Table 1. *Verticillium* isolates used in this study for VCG-related analyses of the IGS subregion P, with code names, VCGs, original hosts, geographic origins (sources), and GenBank accession numbers

Isolate code	VCG*	Original host	Geographic origin (source[†])	GenBank accession number
V. dahliae				
Gh.101	1	Cotton	CA, USA (G)	DQ165203
T9	1A	Cotton	CA, USA (A)	KC152269
V44	1A	Cotton	TX, USA (A)	KC152270
cotVd03	1A	Cotton	Turkey (B)	KC152271
cotVd47	1A	Cotton	Turkey (B)	KC152272
V661I (328 v-1)	1B	Cotton	Greece (C)	KC152273
V666I (347 v-1)	1B	Cotton	Greece (C)	KC152274
Le.112	2	Tomato	CA, USA (G)	FJ931480
Ls.7	2	Lettuce	CA, USA (G)	DQ165207
Cs.80	2	Artichoke	CA, USA (G)	DQ165201
Cv.85	2	Watermelon	CA, USA (G)	DQ165240
PH	2A	Pistachio	CA, USA (A)	KC152275
179-4	2A	Melon	Greece (D)	KC152276
164-1	2A	Tomato	Greece (D)	KC152277
98-1	2A	Tomato	Greece (D)	KC152278
cotVd04	2A	Cotton	Turkey (B)	KC152279
V54	2B	Pepper	Austria (E)	KC152280
115	2B	Cotton	Syria (A)	KC152281
cot274	2B	Cotton	Israel (F)	KC152282
449-2	2B	Broccoli	Greece (C)	KC152283
V38	2B	Sunflower	Germany (E)	KC152284
V49	2B	Pepper	Austria (E)	KC152285
V57	2B	Strawberry	Germany (E)	KC152286
V613I	2B	Artichoke	Spain (C)	KC152287
Ls.17	2B	Lettuce	CA, USA (G)	KC152288
124-8	2B (1)	Tomato	Greece (D)	KC152289
997-1	2B (2A)	Eggplant	Greece (D)	KC152290
150-5	2B (2A, 4A)	Tomato	Greece (D)	KC152291
999-1	2B (4A, 4B)	Eggplant	Greece (D)	KC152292
302-1	2B (2A)	Tomato	Greece (D)	KC152293
530-1	2B (2A)	Pepper	Greece (D)	KC152294
25V (SS4)	2AB (1, 4A)	Cotton	USA (H)	KC152295
70-21	3	Pepper	AZ, USA (A)	KC152296
PCW	3	Pepper	CA, USA (A)	KC152297
Ca.59	4	Bell pepper	CA, USA (G)	DQ165195
Le.78	4	Tomato	CA, USA (G)	FJ931478
Le.88	4	Tomato	CA, USA (G)	FJ931479
Le.109	4	Tomato	CA, USA (G)	DQ165238
Sm.113	4	Eggplant	MD, USA (G)	DQ165205
St.91	4	Potato	ID/OR, USA (G)	DQ165231
BB	4A	Potato	ID, USA (A)	KC152298
30-6	4A	Potato	Canada (I)	KC152299
171-4A	4A	Potato	USA (C)	KC152300
V830	4A	Potato	OH, USA (E)	KC152301
S39	4B	Soil	OH, USA (A)	KC152302 (clone B), KC152303 (clone C)
pn4	4B	Peanut	Israel (F)	KC152304
453-1	4B	Vetch	Greece (D)	KC152305
473-1	4B	Summer squash	Greece (D)	KC152306
V684I	4B	Artichoke	Spain (C)	KC152307
V39	4B	Sunflower	Germany (E)	KC152308
451-1	4B (4A)	Lettuce	Greece (D)	KC152309
465-2	4B (4A)	Marigold	Greece (D)	KC152310
802-1	4B (4A)	Olive	Greece (D)	KC152311
467-2	4B (4A)	White top	Greece (D)	KC152312

Table 1. Continued

Isolate code	VCG*	Original host	Geographic origin (source†)	GenBank accession number
lt1-1	4B (3, 4A)	Lettuce	Greece (D)	KC152313
Ca.83	6	Bell pepper	CA, USA (G)	KC152314
Ca.146	6	Bell pepper	CA, USA (G)	KC152315
Ca.148	6	Bell pepper	CA, USA (G)	KC152316
Cf.38	6	Chili pepper	CA, USA (G)	KC152317
V. albo-atrum				
Ms.102	–	Alfalfa	PA, USA (G)	KC152318
Ms.103	–	Alfalfa	PA, USA (G)	DQ165192
Ms.107	–	Alfalfa	PA, USA (G)	FJ931528
Ms.108	–	Alfalfa	PA, USA (G)	GQ336791
V. longisporum				
Vlon.K12	–	Oilseed rape	Russia (J)	KC152319
Bp.1	–	Napa cabbage	CA, USA (G)	DQ165190
Bob.70	–	Cauliflower	CA, USA (G)	DQ165187
Ar.136	–	Horseradish	IL, USA (G)	DQ165183
Bno.188	–	Oilseed rape	Germany (G)	DQ165185

*For certain *V. dahliae* isolates used in this study, data on 'bridging' behavior (isolates complementing to varying degrees the testers of more than one VCG) were available; these secondary VCG interactions are provided in brackets.

†A = R. Rowe, OARDC, The Ohio State University, USA; B = S. Dervis, University of Mustafa Kemal, Turkey; C = M. Jiménez-Gasco, The Pennsylvania State University, USA; D = E. Ligoxigakis, Plant Protection Institute, N.AG.RE.F., Greece; E = A. von Tiedemann, University of Göttingen, Germany; F = T. Katan, The Volcani Center, Israel; G = K. Subbarao, University of California, Davis, USA; H = E. Paplomatas, Agricultural University of Athens, Greece; I = K. Dobinson, University of Western Ontario, Canada & Agriculture and Agri-Food, Canada; J = O. Strunnikova, All-Russian Research Institute for Agricultural Microbiology, Russia.

period, 15 001 trees were sampled from the posterior probability distribution (every 100 cycles) during the sampling period (1 500 000 generations). Two independent MCMCMC (Metropolis-Coupled Markov Chain Monte Carlo) searches were run using different random starting points.

Results

Structural analysis of the IGS region of *V. dahliae* and related species

The IGS sequences of 201 *V. dahliae* isolates available at the GenBank database were retrieved and aligned (Table S1). A pronounced length variation was observed, as complete sequences ranged in size from 1212 to 1968 bp. Based on the distribution of polymorphisms along the multiple sequence alignment, three distinct subregions were identified: a central, highly polymorphic subregion P (nucleotide positions 361–1189 with regard to the IGS sequence of *V. dahliae* isolate Ls.17; genome sequenced, Broad Institute, at http://www.broadinstitute.org/annotation/genome/verticillium_dahliae/MultiHome.html), flanked by two highly conserved subregions, C1 (nucleotides 1–360) and C2 (nucleotides 1190–1801), as shown in Fig. 1a. Sequence conservation ranged between 95.3% and 100% (pairwise identities) for subregion C1 and 97.8% and 100% for C2. Sequence polymorphism in these two

conserved subregions was due mainly to evenly scattered single-base substitutions (with transitions being twofold more frequent than transversions for both subregions) and, to a lesser extent, few small (mainly 1- to 2-bp) insertions/deletions (indels). Subregion P was less conserved (percentages of pairwise identities as low as 83.0% were identified) and was characterized by a multitude of indels of varying sizes as well as several point mutations.

A structural analysis of the IGS sequences (derived from *V. dahliae* isolates of different original hosts and geographic origins) revealed the ubiquitous presence of four repetitive elements, organized in different patterns as interwoven sets of repeats (Fig. 1b and c). The most commonly encountered element was an 18-bp imperfect repeat (element R) found in all sequences examined, in five to 14 copies. All R copies were direct repeats, apart from the first, imperfect inverted copy (R'), which was located in subregion C1 (nucleotide 266). The first direct R copy (nucleotide 360) was invariably located at the beginning of the polymorphic subregion P, while a pair of successive R copies could be identified in the 3′ region of this subregion (nucleotide 964) in most sequences. Although the majority of R copies were perfectly conserved among all sequences, a number of sequence variants were also identified (Fig. 1d). In addition to the R repeats, three more classes of short repetitive elements with highly conserved sequences were identified, namely A (17 bp; imperfect direct repeats, found in one to four
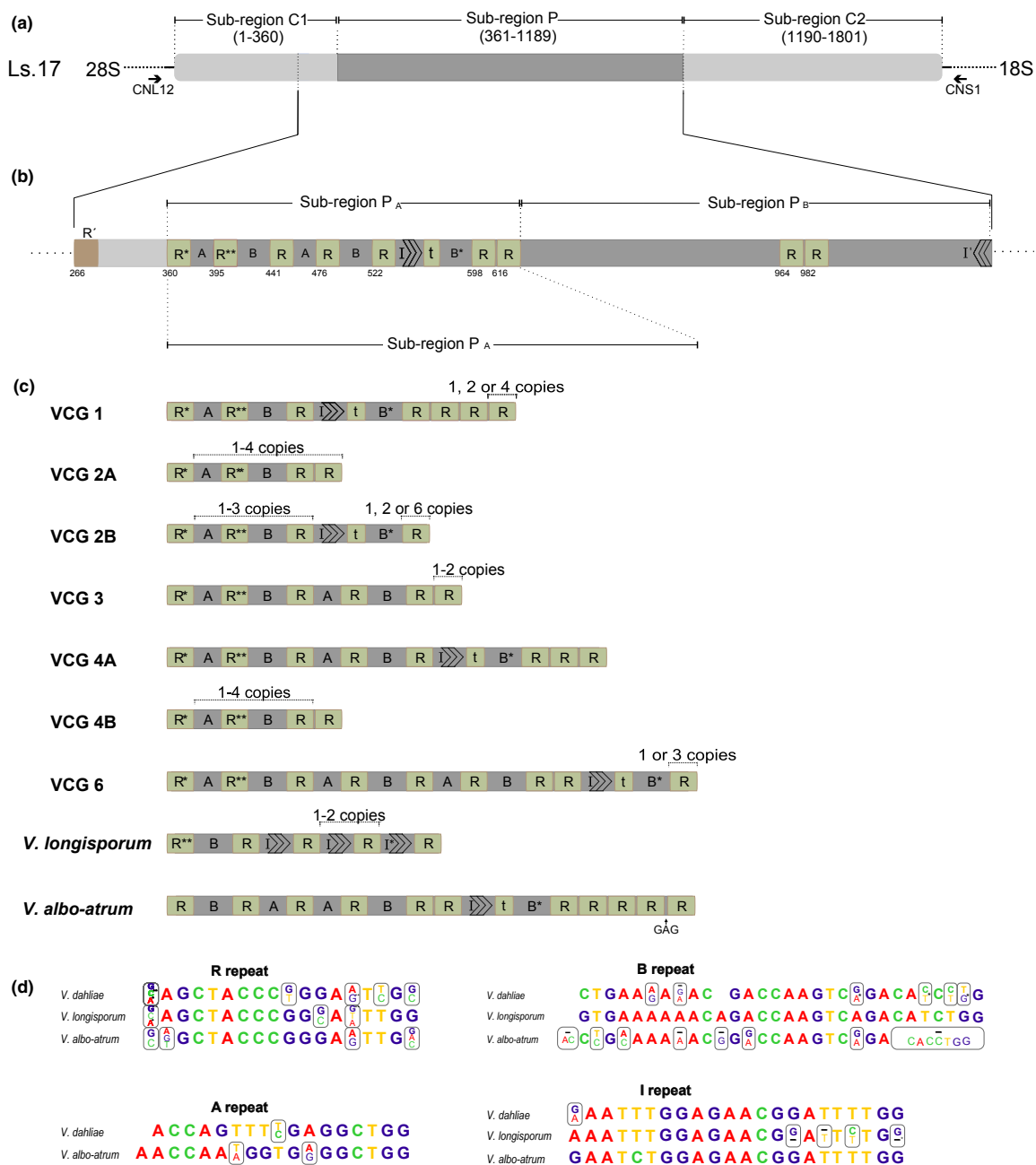
(a)

Sub-region C1 (1-360) Sub-region P (361-1189) Sub-region C2 (1190-1801)

Ls.17 28S ···· CNL12 18S CNS1

(b)

Sub-region P$_A$ Sub-region P$_B$

R′

R* A R** B R A R B R I t B* R R R R I′

266 360 395 441 476 522 598 616 964 982

Sub-region P$_A$

(c)

1, 2 or 4 copies

VCG 1 R* A R** B R I t B* R R R R

1-4 copies

VCG 2A R* A R* B R R

1-3 copies 1, 2 or 6 copies

VCG 2B R* A R** B R I t B* R

1-2 copies

VCG 3 R* A R** B R A R B R R

VCG 4A R* A R** B R A R B R I t B* R R R

1-4 copies

VCG 4B R* A R** B R R

1 or 3 copies

VCG 6 R* A R** B R A R B R A R B R R I t B* R

1-2 copies

V. longisporum R** B R I R I R I R

V. albo-atrum R B R A R A R B R R I t B* R R R R R

G A G

(d)

R repeat
V. dahliae A G C T A C C C G G A A T G G
V. longisporum A G C T A C C C G G G A T T G G
V. albo-atrum G G C T A C C C G G G A A T T G G

B repeat
V. dahliae C T G A A A A A A C G A C C A A G T C G G A C A C C T G
V. longisporum G T G A A A A A A C A G A C C A A G T C A G A C A T C T G G
V. albo-atrum A C C C G A A A A A A C G G C C A A G T C G A C A C C T G G

A repeat
V. dahliae A C C A G T T T C G A G G C T G G
V. albo-atrum A A C C A A G G T G G G G C T G G

I repeat
V. dahliae G A A A T T T G G A G A A C G G A T T T T G G
V. longisporum A A A T T T G G A G A A C G A T T G T G G
V. albo-atrum G A A T C T G G A G A A C G G A T T T T G G

Fig. 1. (a) Diagrammatic representation of the nuclear rDNA IGS region of *Verticillium dahliae* isolate Ls.17, divided into subregions C1, P, and C2, with corresponding nucleotide positions indicated in parentheses. The arrows below the bar indicate the annealing positions of PCR primers CNL12 and CNS1, within the extremities of the 28S and 18S rRNA genes, respectively. (b) Organization of direct and inverted repetitive elements R, A, B, and I in the IGS polymorphic subregion P of *V. dahliae* Ls.17. The division of subregion P into two areas (P$_A$ and P$_B$) is shown above the bar, while nucleotide start positions of R copies are provided below the bar. Asterisks next to the designations of R and B elements indicate sequence variants of them; the corresponding sequences are shown in Fig. 1d. (c) Consensus structural organization of IGS subregion P (area P$_A$) sequences of different *V. dahliae* VCGs and *V. albo-atrum* and *V. longisporum* isolates. The asterisk next to the I element's designation indicates a sequence variant of it; the corresponding sequence is shown in Fig. 1d. Area P$_B$ is not included because it has the same structural organization in all isolates. (d) Sequences of repetitive elements R, A, B, and I and their variants present in isolates of the three *Verticillium* species examined. For each element, the sequences identified in *V. dahliae*, *V. albo-atrum*, and *V. longisporum* isolates are provided. All polymorphic positions are indicated with boxes; in each box, the upper base is the one most frequently encountered in the population, whereas substitutions present in sequence variants in one or more isolates are provided below. Asterisks next to bases in polymorphic sites correspond to variants of the elements as presented in Fig. 1b and c. Single- or multibase deletions are indicated with dashes.

copies), B (28 bp; imperfect direct repeats, one to four copies), and I (21–22 bp; a pair of perfect inverted repeats) (Fig. 1b–d). Elements R, A, and B were ubiquitously found in all isolates tested, whereas element I was identified in most of them. Interestingly, the four classes of repetitive elements were often associated in higher-order sequence repetitions (e.g. the structure R-A-R-B was identified in all sequences, in one to four consecutive iterations, as a large 81-bp repetitive element), or in long conserved composite blocks (e.g. the first I inverted repeat, whenever present, was always followed by a truncated form of the R element (t), then by a B repeat, and by at least one R copy: I-t-B-R) (Fig. 1b and c). The size variation observed among IGS sequences was mainly due to the varying numbers of the four short conserved repetitive elements, within the boundaries of the P subregion. Based on the distribution of such repetitive elements, subregion P could be divided into two distinct characteristic areas, P_A and P_B. Area P_A extended from the first direct R copy (nucleotide 360) to the last block of R copies in a successive row of repetitive elements (nucleotide 634), and area P_B followed P_A as far as the 3′ end of subregion P (Fig. 1b and c). Area P_B contained only a pair of successive R copies (nucleotide 964) and the second I inverted repeat at its 3′ end, thus being the most conserved DNA stretch within subregion P.

When the IGS sequences of four representative *V. albo-atrum* and five *V. longisporum* isolates were added to the analysis (Table 1), it became evident that the three species shared a common structural organization (Fig. 1c), consisting of variants of the same repetitive elements described for *V. dahliae* (Fig. 1d). Elements R and I were well conserved, while A and B repeats were more divergent. In comparison with *V. dahliae*, *V. albo-atrum* isolates had a higher level of R element repetition in subregion P, with 14 direct almost perfect copies and another truncated iteration. All *V. longisporum* strains shared a unique organization of the same subregion, although isolated from different hosts and distant geographic origins. The first conserved direct R copy was followed by a sequence variant of element B, to be then followed by three to four iterations of an R-I motif (Fig. 1c). The rest of the elements, that is, the conserved pair of R direct repeats in the 3′ region of the subregion P and the downstream inverted copy of element I, were organized in the same way as in *V. dahliae* isolates.

Structural and phylogenetic analysis of IGS subregion P sequences among VCGs of *V. dahliae*

The IGS subregion P sequences of all isolates listed in Table 1 (covering all *V. dahliae* VCGs) were obtained, structurally analyzed, and used for phylogenetic analysis. Isolates within each VCG subgroup were structurally similar, while characteristic differences were observed among different VCG subgroups (Fig. 1c). Isolates belonging to VCGs 2A, 4B, and 3 shared a common organization of R-A-R-B iterations in their IGS sequences. On the other hand, isolates from VCGs 1, 2B, 4A, and 6 were characterized by an additional I-t-B*-R block (with B* corresponding to a particular sequence variant of element B; Fig. 1d). A $R_{(4–8)}$ box was unique to all members of VCG 1. The only apparent contradiction to this classification was isolate V39 with an I-t-B*-R block (albeit VCG 4B). VCG 2B was the most polymorphic VCG subgroup, with isolates 530-1, V613I, and V49 exhibiting unusual copy numbers of the R element or the R-A-R-B structure.

The phylogeny produced with IGS subregion P sequence analysis was generally consistent with the structural study. Two main intraspecific clades were distinguished (I and II) with excellent phylogenetic support, with the former including isolates from VCGs 2A, 4B, 3, and 1 and the latter consisting of isolates from VCGs 2B, 4A, and 6 (Fig. 2). Clade I was further unambiguously subdivided into two subclades, one of them encompassing all VCG 1 isolates (of both subgroups 1A and 1B), distinctly from VCGs 2A, 4B, and 3 (grouped in the other subclade). Isolate V39 (VCG 4B), which was structurally different in the IGS subregion P from the other VCG 4B strains, grouped with VCG 1 isolates rather than the 2A/4B subclade. The two isolates of VCG 3 were placed within the 2A/4B subclade. In clade II, a trend of distinction among VCGs 2B, 4A, and 6 was also observed, although it was only moderately supported by phylogenetic analysis (Fig. 2). VCG 2B was the most heterogeneous group, with the structurally aberrant isolates 530-1 and V613I grouping with VCG 1, in clade I. To exclude the possibility of previous misclassification of these exceptional isolates in VCGs (V39, 530-1, and V613I), we repeated their VCG assignment using international tester strains, and their original classification was indeed verified (data not shown).

Fig. 2. Phylogenetic tree derived from ML analysis of the IGS subregion P sequences (based on a sequence dataset of 1144 bp including gaps) of 68 *Verticillium* isolates (MEGA 5.05 software). ML bootstrap support values above 50% (upper numbers) calculated from 1000 replicates and BI posterior probabilities (lower numbers) produced over 1 500 000 generations (MRBAYES 3.2.1) are shown at nodes. Next to each isolate's code, the VCG classification of the isolate is provided in square brackets or in parentheses for those 'bridging' isolates reacting with the testers of more than one VCG (Table 1). On the right of the figure, the grouping of isolates according to VCGs and main clades (for *V. dahliae*) or species (for *V. albo-atrum* and *V. longisporum*) is indicated.
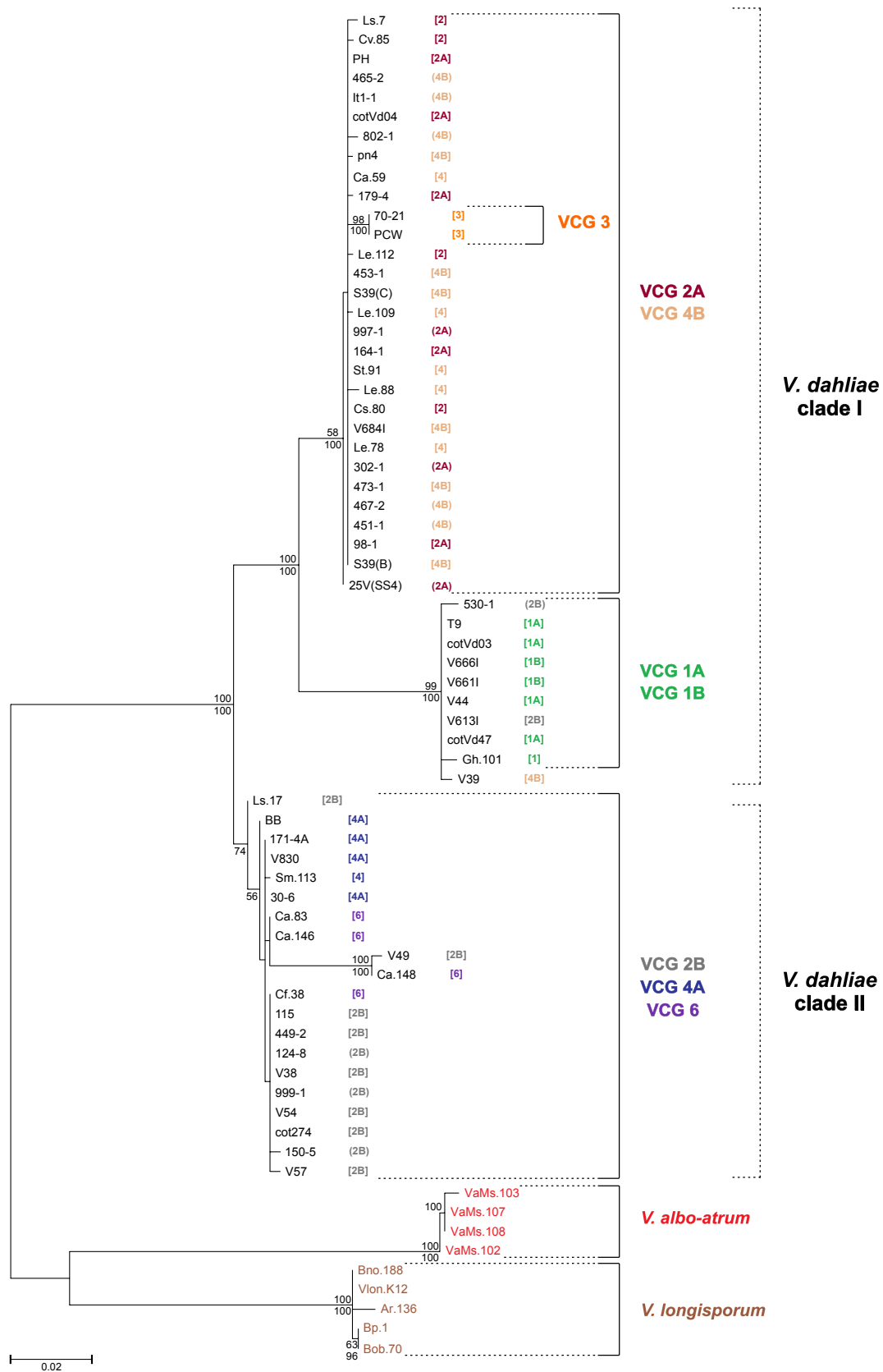
Two IGS length variants, clones S39(B) and S39(C), were detected in the case of isolate S39 (VCG 4B) during our cloning experiments, clearly indicating heterogeneity of the IGS region in this isolate. The sequences were identical except for the presence of an additional R-A-R-B box in the larger variant (C). The grouping of both sequences in the phylogenetic tree was the same, together with the majority of VCG 4B isolates. Ten of the *V. dahliae* isolates tested for which only main VCG assignments were available (i.e. VCGs 2 and 4, but unknown VCG subgroups; Table 1) were all unambiguously placed in tree clades that were associated with specific subgroups of these VCGs. Similarly, for twelve of the isolates analyzed, which are known to establish heterokaryons with the testers of more than one VCG subgroup ('bridging' strains; Table 1), grouping with only one of their compatible VCG clades was achieved. Overall, the IGS phylogeny-derived grouping of *V. dahliae* isolates was irrespective of host source and geographic origin. Finally, when four *V. albo-atrum* and five *V. longisporum* isolates were added to the analysis, excellent discrimination was achieved among the three plant pathogenic *Verticillium* species (Fig. 2).

Discussion

Comparative analysis of the nuclear rDNA IGS region of *V. dahliae* allowed the identification of a highly polymorphic subregion (P), flanked by two more conserved areas (C1 and C2). This organization putatively reflects the different functioning of these subregions, with the highly conserved subregions probably accommodating functions related to rRNA production and processing, while the most polymorphic subregion P might be responsible for the promotion of unequal crossing-over events and the maintenance of homogeneity between rDNA complexes (Pramateftaki *et al.*, 2000; Ganley & Kobayashi, 2011). Novel alterations in the IGS region are predicted to spread to neighboring spacer regions through the mechanisms of concerted evolution, by means of unequal chromatid exchange and biased gene conversion (Eickbush & Eickbush, 2007). A prediction of this model would be the presence of IGS variants in the same nucleus, at detectable frequencies, corresponding to intermediates in the process of slow homogenizing concerted evolution (Ganley & Scott, 2002). Indeed, although only five clones were checked for insert presence prior to sequencing, two length variants were detected in the case of isolate S39, differing from each other in the number of R-A-R-B iterations. It is, thus, hypothesized that this conserved repetitive box participates in the generation of polymorphism in the IGS region. The four classes of repetitive elements that were

identified, ubiquitous, and conserved among the three *Verticillium* species tested may constitute key factors of this mechanism. The extended heterogeneity in the IGS of *V. dahliae* should be taken into consideration when this region is used for the development of assays for detection and quantification (Bilodeau *et al.*, 2012), as it might reduce the accuracy of such practical tools.

Structural and phylogenetic analyses of the polymorphic IGS subregion P sequences confirmed that isolates within VCGs are molecularly similar, regardless of original host and geographic origin, in agreement with Collado-Romero *et al.* (2006). A clear division of *V. dahliae* into two intraspecific lineages was demonstrated, with clade I encompassing VCGs 1, 2A, 4B, and 3, whereas clade II included VCGs 2B, 4A, and 6. Remarkably, this grouping of VCGs was essentially the same as these produced before with multilocus sequencing, AFLP fingerprinting, and mitochondrial haplotype analyses (Collado-Romero *et al.*, 2008; Martin, 2010). Therefore, it was verified that even though it is based on a single DNA locus, the IGS polymorphic area-based method that we tested here produces a highly congruent phylogenetic profile with that known from previous comprehensive analyses of the fungus and can be used to reliably depict phylogenetic relations among VCGs. Moreover, because analysis of the whole IGS region of representative VCG tester strains led to an identical phylogeny (data not shown), sequencing of only the polymorphic subregion P significantly reduces the required amount of sequencing needed for an extended fungal population and, thus, is a practical tool for high-throughput population profiling purposes. An exception to this general congruence between the different methods was VCG 1, which was grouped externally to the rest of the intraspecific phylogeny in previous studies. A future study of *V. dahliae* VCGs with a more diverse collection of VCG 1 isolates may clarify the genetic relationships of VCG 1 with other VCGs, as well as distinguish between its subgroups, which remained undifferentiated with IGS analysis conducted here.

Isolates from VCGs 2A and 4B structurally and phylogenetically grouped together (in clade I) and distinctly from their 'sister' VCG subgroups 2B and 4A, respectively. The latter subgroups were also closely related to each other (in clade II). The clear differentiation of VCG 2A from 2B and VCG 4A from 4B is in agreement with previous observations (Collado-Romero *et al.*, 2006, 2008; Martin, 2010; Papaioannou *et al.*, 2013). Thus, based on the genetic relationships inferred from this pattern, it is proposed that main VCGs 2 and 4 are not genetically meaningful, and their use in future VCG classification studies should be reconsidered. Similarly, previous observations of cross-reactions of VCG 3 isolates with subgroups of VCG 4 (Strausbaugh *et al.*, 1992), which

possibly account for the rare report of such isolates from populations of the fungus and their omission from most phylogenetic studies, together with our finding that the only two available representative isolates were not differentiated from the VCG 2A/4B subclade, suggest altogether that this VCG may actually be an artifact of the methodology used for vegetative compatibility grouping and, thus, of limited utility in the population genetics of *V. dahliae*. Furthermore, Collado-Romero *et al.* (2008) previously concluded that VCG 2B is polyphyletic, based on multilocus and AFLP fingerprinting analyses. In agreement with that study, some heterogeneity was also observed in the collection tested here for VCG 2B as well as, to a lesser extent, within VCG 4B, with aberrant isolates grouping in the VCG 1-related subclade (clade I). Finally, the three VCG 6 isolates from bell pepper (Ca.83, Ca.146, and Ca.148) grouped together within cluster II, whereas the sole chili pepper VCG 6 isolate (Cf.38) was placed with the majority of VCG 2B isolates in the same cluster. This may be attributed to some restricted host-related isolation, as strains from the two pepper varieties have also been shown to differ in other molecular characteristics (I.A. Papaioannou & M.A. Typas, unpublished data).

In addition to its contribution to the phylogenetic study of VCGs, the IGS analysis has great potential to facilitate intra- and interspecific differentiation of *Verticillium* species. Within *V. dahliae*, we showed that IGS analysis is very useful for confirmation/disambiguation (e.g. for VCG 'bridging' isolates), for prediction (e.g. for isolates that have been only partially VCG assigned), and thus for profiling of *V. dahliae* VCG classification on large-scale analyses. Moreover, the clear discrimination achieved among the three plant pathogenic *Verticillium* species with sequence analysis of only one genetic locus (IGS subregion P) renders this method a suitable candidate for rapid and accurate interspecific discrimination.

Acknowledgements

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Atallah ZK, Maruthachalam K, du Toit L, Koike ST, Davis RM, Klosterman SJ, Hayes RJ & Subbarao KV (2010) Population analyses of the vascular plant pathogen *Verticillium dahliae* detect recombination and transcontinental gene flow. *Fungal Genet Biol* **47**: 416–422.

Bhat RG, Smith RF, Koike ST, Wu BM & Subbarao KV (2003) Characterization of *Verticillium dahliae* isolates and wilt epidemics of pepper. *Plant Dis* **87**: 789–797.

Bilodeau GJ, Koike ST, Uribe P & Martin FN (2012) Development of an assay for rapid detection and quantification of *Verticillium dahliae* in soil. *Phytopathology* **102**: 331–343.

Collado-Romero M, Mercado-Blanco J, Olivares-García C, Valverde-Corredor A & Jiménez-Díaz RM (2006) Molecular variability within and among *Verticillium dahliae* vegetative compatibility groups determined by fluorescent amplified fragment length polymorphism and polymerase chain reaction markers. *Phytopathology* **96**: 485–495.

Collado-Romero M, Mercado-Blanco J, Olivares-García C & Jiménez-Díaz RM (2008) Phylogenetic analysis of *Verticillium dahliae* vegetative compatibility groups. *Phytopathology* **98**: 1019–1028.

Collado-Romero M, Berbegal M, Jiménez-Díaz RM, Armengol J & Mercado-Blanco J (2009) A PCR-based "molecular tool box" for *in planta* differential detection of *Verticillium dahliae* vegetative compatibility groups infecting artichoke. *Plant Pathol* **58**: 515–526.

Daayf F, Nicole M & Geiger J-P (1995) Differentiation of *Verticillium dahliae* populations on the basis of vegetative compatibility and pathogenicity on cotton. *Eur J Plant Pathol* **101**: 69–79.

Eickbush TH & Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**: 477–485.

Ganley ARD & Kobayashi T (2011) Monitoring the rate and dynamics of concerted evolution in the ribosomal DNA repeats of *Saccharomyces cerevisiae* using experimental evolution. *Mol Biol Evol* **28**: 2883–2891.

Ganley ARD & Scott B (2002) Concerted evolution in the ribosomal RNA genes of an Epichloë endophyte hybrid: comparison between tandemly arranged rDNA and dispersed 5S rrn genes. *Fungal Genet Biol* **35**: 39–51.

Hong SG, Liu D & Pryor BM (2005) Restriction mapping of the IGS region in *Alternaria* spp. reveals variable and conserved domains. *Mycol Res* **109**: 87–95.

Jiménez-Díaz RM, Mercado-Blanco J, Olivares-García C, Collado-Romero M, Bejarano-Alcázar J, Rodríguez-Jurado D, Giménez-Jaime A, García-Jiménez J & Armengol J (2006) Genetic and virulence diversity in *Verticillium dahliae* populations infecting artichoke in eastern-central Spain. *Phytopathology* **96**: 288–298.

Joaquim TR & Rowe RC (1990) Reassessment of vegetative compatibility relationships among strains of *Verticillium*

dahliae using nitrate-nonutilizing mutants. *Phytopathology* **80**: 1160–1166.

Katan T (2000) Vegetative compatibility in populations of Verticillium – an overview. Advances in Verticillium: research and disease management. *Proceedings of the 7th International Verticillium Symposium* (Tjamos EC, Rowe RC, Heale JB & Fravel DR, eds), pp. 69–86. The American Phytopathological Society, St. Paul, MN.

Leslie JF (1993) Fungal vegetative compatibility. *Annu Rev Phytopathol* **31**: 127–150.

Martin FN (2010) Mitochondrial haplotype analysis as a tool for differentiating isolates of *Verticillium dahliae*. *Phytopathology* **100**: 1231–1239.

Mbofung GY, Hong SG & Pryor BM (2007) Phylogeny of *Fusarium oxysporum* f. sp. *lactucae* inferred from mitochondrial small subunit, elongation factor 1-α, and nuclear ribosomal intergenic spacer sequence data. *Phytopathology* **97**: 87–98.

Morton A, Tabrett AM, Carder JH & Barbara DJ (1995) Sub-repeat sequences in the ribosomal RNA intergenic regions of *Verticillium albo-atrum* and *V. dahliae*. *Mycol Res* **99**: 257–266.

Pantou MP, Mavridou A & Typas MA (2003) IGS sequence variation, group-I introns and the complete nuclear ribosomal DNA of the entomopathogenic fungus *Metarhizium*: excellent tools for isolate detection and phylogenetic analysis. *Fungal Genet Biol* **38**: 159–174.

Papaioannou IA, Ligoxigakis EK, Vakalounakis DJ, Markakis EA & Typas MA (2013) Phytopathogenic, morphological, genetic and molecular characterization of a *Verticillium dahliae* population from Crete, Greece. *Eur J Plant Pathol* **136**: 577–596.

Pegg GF & Brady BL (2002) *Verticillium Wilts*. CABI Publishing, New York, NY.

Pramateftaki PV, Antoniou PP & Typas MA (2000) The complete DNA sequence of the nuclear RNA gene complex of *Verticillium dahliae*: intraspecific heterogeneity within the intergenic spacer region. *Fungal Genet Biol* **29**: 19–27.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA & Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**: 539–542.

Rowe RC (1995) Recent progress in understanding relationships between *Verticillium* species and subspecific groups. *Phytoparasitica* **23**: 31–38.

Strausbaugh CA, Schroth MN, Weinhold AR & Hancock JG (1992) Assessment of vegetative compatibility of *Verticillium dahliae* tester strains and isolates from California potatoes. *Phytopathology* **82**: 61–68.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Tran VT, Braus-Stromeyer SA, Timpner C & Braus GH (2013) Molecular diagnosis to discriminate pathogen and apathogen species of the hybrid *Verticillium longisporum* on the oilseed crop *Brassica napus*. *Appl Microbiol Biotechnol* **97**: 4467–4483.

Typas MA, Griffen AM, Bainbridge BW & Heale JB (1992) Restriction fragment length polymorphisms in mitochondrial DNA and ribosomal RNA gene complexes as an aid to the characterization of species and sub-species populations in the genus *Verticillium*. *FEMS Microbiol Lett* **95**: 157–162.

White TJ, Bruns TD, Lee SB & Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications* (Innis MA, Gelfand DH, Sninsky JJ & White TJ, eds), pp. 315–322. Academic Press, New York, NY.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Sequences of the IGS region of *V. dahliae* isolates used for structural analyses in this study, with code names, original hosts, geographic origins, sizes of the IGS region and GenBank accession numbers.