Mycopathologia (2008) 166:1–16
DOI 10.1007/s11046-008-9094-3

Domain II Hairpin Structure in ITS1 Sequences as an Aid in Differentiating Recently Evolved Animal and Plant Pathogenic Fungi

P. D. Bridge · T. Schlitt · P. F. Cannon · A. G. Buddie · M. Baker · A. M. Borman

Received: 9 July 2007 / Accepted: 4 February 2008 / Published online: 14 March 2008

Abstract The hypothesis that ITS structural features can be used to define fungal groups, where sequence analysis is unsatisfactory, was examined in plant and animal pathogenic fungi. Structural models of ITS1 regions were predicted for presumed closely related species in *Colletotrichum* and *Trichophyton* anamorphs of *Arthroderma* species. Structural alignment of models and comparison with ITS sequence analysis identified a variable region in a conserved hairpin formed from a common inverted repeat. Thirteen different hairpin structure models were obtained for *Colletotrichum* species and five different models were obtained for *Trichophyton* species. The different structure types could be matched to individual species and species complexes as defined by ITS sequence analysis.

Keywords *Colletotrichum* · *Trichophyton* · *Arthroderma* · ITS · Structural models · Sequence analysis

P. D. Bridge (✉) · T. Schlitt
Biological Sciences Division, British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK
e-mail: pdbr@bas.ac.uk

P. F. Cannon · A. G. Buddie
CABI Europe – UK, Bakeham Lane, Egham, Surrey TW20 9TY, UK

M. Baker
East Kent Microbiology Service, The William Harvey Hospital, Kennington Road, Ashford, Kent TN24 0LZ, UK

A. M. Borman
Mycology Reference Laboratory, Health Protection Agency South-West Regional Laboratory, Myrtle Road, Bristol BS2 8EL, UK

Introduction

Plant and animal pathogenic fungi can show considerable host specificity and can be geographically or ecologically isolated. Increased travel and movement of materials, together with plant and animal breeding programmes, usage of pesticides and antimicrobials, and farming and social conditions are all examples of anthropogenic change that can affect the distribution, activity and specificity of fungal pathogens. DNA-based systematics generally reflect evolutionary age, with characteristics of conserved regions such as protein coding genes being used to identify deep phylogenetic divisions such as family and order, while more variable regions such as spacers have been used to differentiate more recent change at a species or subspecies level [1]. In cases where there has been very recent host speciation, such as in pathogens of cultivated plants or emerging animal disease, very recent evolutionary events may not be apparent in selected individual DNA sequences. In this situation, broader total genome approaches such as AFLP may be necessary to identify individual taxa [2, 3].

The use of sequences from internal transcribed spacer (ITS) of the ribosomal RNA (rRNA) gene cluster has become established as a primary criterion for identifying and classifying fungal pathogens at the species level (e.g. [4–6]). Although clear differences in sequence can be diagnostic, the interpretation of very low levels of sequence variation is more problematic [1]. This is particularly the case in presumably relatively recently evolved groups, such as plant pathogens of cultivated crops and emerging human and animal pathogens [7]. In such cases, total ITS sequence differences may be limited to only a few base changes or short deletions, and it can be difficult to determine the relative significance of such changes in defining the taxa, sometimes leading to claims of "over classification" [8].

The rRNA ITS region is transcribed, but it is not translated and does not give rise to a gene product. It is an essential region and removal of the various domains has different effects on the production of the rRNA subunits [9]. The current model for rRNA formation involves the transcription of the complete rRNA gene cluster as a single unit and subsequent folding of the RNA prior to translation (see [10]). Particular structures in the ITS transcript, such as helices and loops, are believed to play a role in anchoring and folding the complete RNA molecule, and to act as markers to ensure their exclusion from the final rRNA–protein complexes [9, 11].

Potentially structural features of ITS1 regions have been investigated for some plant and animal groups (e.g [10, 12, 13]). Early studies on ribosomal gene structure were developed from ascomycetous yeast models, and conserved hairpin structures have been identified in their ITS sequences [9]. Phylogenetic signals have been reported from structural models of variable regions in basidiomycete mitochondrial rRNA genes [14] and sequence and secondary structure motifs associated with the variable regions of ITS1 have been used to identify genera of rumen inhabiting chytrid fungi [15]. More recently, potentially species-specific structures have been identified in some fungal ITS2 regions [16].

The ascomycetes *Colletotrichum* Corda and *Trichophyton* Malmsten are good candidates for investigating close taxonomic relationships. They are both genera that have clear links to known human history, as pathogens of crop plants and humans, respectively. In both genera, it can be argued that speciation has occurred relatively recently, and in both genera there are well established groups of species considered to be species complexes (see [8, 17]). There has been extensive use of the ITS sequences in these genera for systematics and identification, and this work suggests that ITS sequences can be used to define clear species and species complex groups, although their use for defining inter-species relationships is less clear (e.g. [8, 18–20]). In both genera the ITS variation is quite limited; in *Colletotrichum* inter-species ITS1 variation is commonly 10–15% and this is largely located in one region, whereas in *Trichophyton* inter-specific variation can in some instances be as low as 3–4% [7, 20]. The two genera are taxonomically only distantly related with *Colletotrichum* being placed in the Sordariomycetidae, and *Trichophyton* anamorphs of *Arthroderma* placed in the Onygenales [21].

The aim of this study was to determine if ITS structural features can be used to define fungal groups where sequence analysis is unsatisfactory, poorly supported or lacking in resolution, such as in some relatively recently evolved fungal pathogens.

Methods

Sequences

Sequences of complete ITS1 regions, together with short sections of the small subunit and 5.8S sequences were obtained for the species complexes of *C. gloeosporioides* and *T. terrestre* and other representatives of the genera *Colletotrichum* and *Trichophyton* from EMBL and works in progress (Table 1). Within each genus, sequences were initially aligned with ClustalW [22]. The alignments were checked manually and the ITS1 regions were identified from the annotations available for *C. gloeosporiodes* AY376534 and *T. verrucosum* Z98002.

Secondary Structure and Prediction

The most probable secondary structures for RNA transcripts of the ITS1 sequences were produced using MFold [23] at http://www.bioinfo.rpi.edu/applications/mfold/old/rna/. ITS1 sequences were

Table 1 Sequences used and type of ITS1 conserved feature obtained (see Figs. 1 and 4)

Sequence	Name as deposited	ITS structure type
AB105971	*Glomerella cingulata*	1
AF207792	*Colletotrichum gloeosporioides*	1
AY266395	*Glomerella cingulata*	1
AY245021	*Glomerella cingulata*	1
AF451905	*Glomerella cingulata*	1
AF534466	*Glomerella cingulata*	1
AF272779	*Glomerella cingulata*	1
AB087219	*Glomerella cingulata*	1
AY438549	*Colletotrichum gloeosporioides*	1
AF521198	*Glomerella cingulata*	1
AJ301912	*Colletotrichum fragariae*	1
AY376540	*Colletotrichum kahawae*	1
AY376536	*Colletotrichum gloeosporioides*	1
AJ301986	*Colletotrichum gloeosporioides* f.sp. *aeschynomene*	2
AJ536229	*Colletotrichum gloeosporioides*	2
AY423476	*Glomerella cingulata*	2
AY376532	*Colletotrichum gloeosporioides*	2
AF090855	*Colletotrichum gloeosporioides*	2
AJ311884	*Colletotrichum gloeosporioides*	2
AF488777	*Glomerella cingulata*	2
AF489568	*Glomerella cingulata*	2
AJ311882	*Colletotrichum gloeosporioides*	2
AJ301988	*Colletotrichum gloeosporioides*	2
AY438550	*Colletotrichum gloeosporioides*	2
AY438545	*Glomerella cingulata*	2
AY438546	*Glomerella cingulata*	2
AY438551	*Glomerella cingulata*	2
AY438548	*Glomerella cingulata*	2
AJ301929	*Colletotrichum musae*	2
AJ301904	*Colletotrichum musae*	2
AY376544	*Glomerella cingulata*	2
AY376534	*Colletotrichum gloeosporioides*	2
AY376529	*Colletotrichum crassipes*	3
AJ536231	*Colletotrichum falcatum*	4
AY376527	*Colletotrichum caudatum*	5
AY376539	*Colletotrichum graminicola*	6
AY376542	*Colletotrichum sublineolum*	7
AB105959	*Colletotrichum destructivum*	8
AJ301938	*Colletotrichum fuscum*	8
AB105955	*Colletotrichum higginsianum*	8
AY301955	*Colletotrichum circinans*	9
AY376528	*Colletotrichum coccodes*	9
AY376510	*Colletotrichum acutatum*	10
AJ301958	*Glomerella lindemuthiana*	11

Table 1 continued

Sequence	Name as deposited		ITS structure type
AY376541	*Colletotrichum orbiculare*		11
AJ301941	*Colletotrichum trifolii*		11
AY376521	*Colletotrichum boninense*		12
AY376526	*Colletotrichum capsici*		13
AY376531	*Colletotrichum dematium*		13
AY376543	*Colletotrichum truncatum*		13
Z98002	*Trichophyton verrucosum*		1
AB058851	*Trichophyton verrucosum*		1
Z98004	*Trichophyton verrucosum*		1
AB078898	*Arthroderma benhamiae*		1
AB100263	*Arthroderma benhamiae*		1
AB088676	*Arthroderma benhamiae*		1
AB088678	*Arthroderma benhamiae*		1
AJ270808	*Trichophyton rubrum*		2
AF170474	*Trichophyton soudanense*		2
AF170473	*Trichophyton soudanense*		2
U18352	*Trichophyton rubrum*		2
AY283675	*Trichophyton rubrum*		2
AJ270809	*Trichophyton soudanense*		2
Z97994	*Trichophyton megninii*		2
AJ270807	*Trichophyton rubrum*		2
AF170464	*Trichophyton megninii*		2
AF170454	*Arthroderma simii*		3
AF170455	*Arthroderma simii*		3
AF170475	*Arthroderma simii*		3
AF170478	*Trichophyton tonsurans*		4
AB094064	*Trichophyton tonsurans*		4
Z98006	*Trichophyton tonsurans*		4
Z98008	*Trichophyton tonsurans*		4
Z98001	*Trichophyton metagrophytes*		4
Z97998	*Trichophyton metagrophytes*		4
NCPF362	*Trichophyton terrestre*	Anamorph *of Arthroderma quadrifidum*	5
NCPF467	*Trichophyton terrestre*	Anamorph *of Arthroderma lenticularum*	5a
NCPF469	*Trichophyton terrestre*	Anamorph *of Arthroderma insingulare*	Single
NCPF602	*Trichophyton terrestre*		Single
Outgroup sequences			
L36640	*Plectosphaerella cucumerina*		N/A
AJ000617	*Microsporum canis*		N/A

folded as a single molecule, and with an additional 10–15 bp of the flanking subunit genes to provide a conserved stem [13]. Folding conditions were the default settings for a linear molecule folded at 37°C in 1 M NaCl with no divalent ions. Maximum interior loop size and asymmetry were set at 30, and suboptimality was set between 3 and 5% to give a minimum of three possible structures for each sequence. Structures were visualised as circle graphs and 2-D structural models in the MFold structure viewer.

Identification of Conserved and Variable Regions

Structural features (loops, stems etc.) that were present in >80% of all possible predicted models were considered to be conserved. The sequences of these features were identified from circle graphs in the MFold viewer. The selected regions of the ITS sequences for each genus were aligned in ClustalW and conserved and variable regions were identified manually from comparison of the alignment plots, and through the conservation function of JalView [24]; see http://www.ebi.ac.uk/jalview/documentation.html#conservation. Variable regions of ITS1 were screened against the EMBL data base at EBI with the BLAST2-WU facility (http://www.ebi.ac.uk/blast2/index.html).

Comparison of Secondary Features

The conserved structures were initially sorted manually to identify any obvious groupings. In instances where only minor differences were observed (such as *Colletotrichum* structure groups 1 & 3), all other hairpins and features in the models were also considered. Where these also showed consistent supporting differences, the structure types were considered different. All of the ITS1 conserved/variable regions making up the hairpin structure were aligned. Where the variable regions differed, the sequence and base pairing annotations were extracted from MFold in Vienna 1.6.1 format (http://www.tbi.univie.ac.at/%7Eivo/RNA/). The structures of these regions were then compared in RNAforester (http://bibiserv.techfak.uni-bielefeld.de/rnaforester). This package compares closed RNA secondary structures by calculating similarities from a global alignment. The similarity scores were set to the standard defaults of 1, 0, −10, 10 & 5 for base matches, base mismatches, base deletion, base pair deletion and base pair bond deletion, respectively. Structure trees were obtained by multiple global alignment based on the tree alignment model [25, 26]. This method includes comparisons of free energy and co-variance to predict a consensus structure, and base pairs in the consensus structure below $P = 0.5$ and base pairs below $P = 0.9$ were pruned during the alignment (for details of default and standard settings see http://bibiserv.techfak.uni-bielefeld.de/rnaforester/submission.html). RNAforester constructs a multiple structure alignment in a progressive way, much like ClustalW. Nearest neighbours are joined first, while the last join produces the complete alignment and can be seen as the root of a clustering tree.

Tree Construction

The aligned ITS1 sequences for each genus were analysed through Phylip [27]. In the case of the *Colletotrichum* data set, these include 15 bp from the 5.8S, in order to anchor and correctly align the outgroup. Distance trees were constructed using the Kimura 2 parameter measure and neighbour joining. Additional analyses were by maximum likelihood. Maximum likelihood trees were produced from assuming a constant single sequence variation rate and a transition/transversion rate of 2. The data set was re-sampled by bootstrapping for 100 replicates, jumbled 5 times in the ML analyses, and consensus trees were drawn in Treeview [28]. The trees were rooted with sequences from *Plectosphaerella cucumerina* for the *Colletotrichum* sequences, and *Microsporum canis* for the *Arthroderma* sequences.

Results

Colletotrichum Sequences

All predicted models of the ITS1 RNA transcript from *Colletotrichum* sequences showed a conserved structure between approximately base 50 and base 95 depending on sequence. This feature was a double hairpin structure arising from a loop on a stem. The detail of this feature varied between species, and these variations were also conserved among the different possible models for each sequence. Thirteen different variants were obtained, and these corresponded to either individual species or groups of closely related species (Fig. 1, Table 1). Sequences from isolates of the *C. gloeosporioides* complex folded to give 1 of 2 different structures (see Fig. 1). The sequences of the double hairpins were aligned and compared for the different structures obtained (Fig. 2). The hairpin was comprised of a very variable G/C rich region (75–92%G+C) that included some simple repeats and long stretches of single bases. The variable region was flanked by conserved

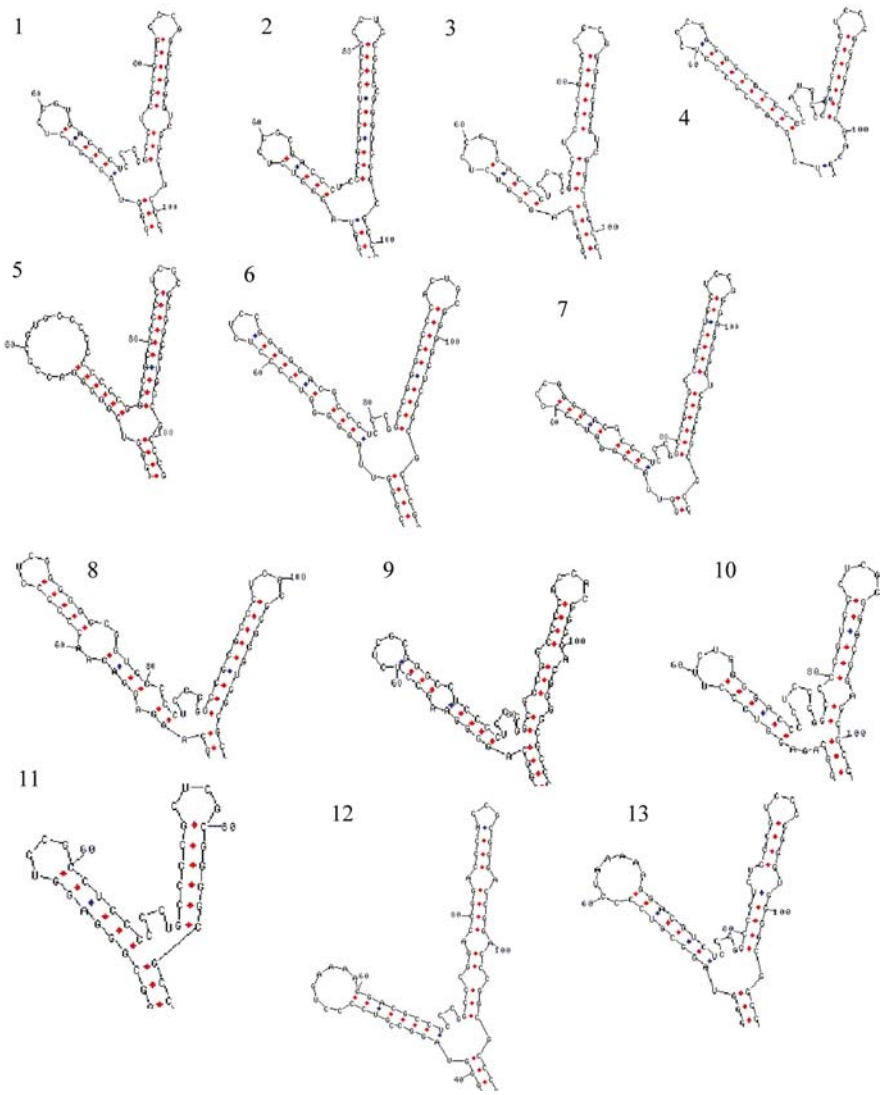


Fig. 1 Thirteen structural variants in *Colletotrichum* ITS1

sequences and anchored by a short inverted repeat of CGGCGG/CCGCCG located at positions 44 and 105 (with reference to AJ376540). The variable sequence within the hairpin structure was not entirely conserved for each structural variant, and the 13 structural variants contained 19 sequence variants, although there were single motifs that could be used to differentiate each structural type (see Fig. 2).

The structures formed by each of the 19 variable region sequences (including the terminal repeats) were compared directly through RNAForester (Fig. 3). In this representation, the 13 structures were aligned as a bifurcating tree. Structure type 11 was recovered in one half of the tree, together with structure types 1–3. Structure types 4–10 and 12 & 13 were recovered in the other half. In all cases, sequence variants of each structure type were recovered together.

The consensus neighbour joining and maximum likelihood trees recovered the sequences in two major clusters (Fig. 4). Bootstrap support for the groupings was somewhat mixed. In general high bootstrap support was obtained for small species groups or individual structure types with the exception of the *C. gloeosporioides* complex. However, there was very little support for subdivisions within these groups or for relationships between them.

The largest cluster contained all of the isolates of the *C. gloeosporioides* complex, and included *C. crassipes* (structure types 1, 2 & 3), Structure

```
Structure
type    Sequence
  1     AY376540  AACTGTTGCTTCGGCGGGT-AGG--GTCTCC-------GTGAC--CCTCCCGGCCT---CCCG--CCCCC--GGG--CGGGTC-GGCGCCCGCCGG
  1     AJ301912  AACTGTTGCTTCGGCGGGT-AGG--GTCCCC-------GTGAC--CCTCCCGGCCT---CCCG--CCCCCCGGG--CGGGTC-GGCGCCCGCCGG
  2     AJ301929  AACTGTTGCTTCGGCGGGT-AGG--GTCCCC-------GTGAC--CCTCCCGGCCC---CCCG--CCCCC--GGG--CGGGTC-GGCGCCCGCCGG
  2     AY376532  AACTGTTGCTTCGGCGGGT-AGG--GTCTCC-------GCGAC--CCTCCCGGCCT---CCCG--CCTCC--GGG--CGGGTC-GGCGCCCGCCGG
  3     AY376529  AACTGTTGCTTCGGCGGGC-AGG--GTCTCC-------GTGAC--CCTCCCGGCCT---CCCG--CCCCC--GGG--CGGGTC-GGCGCCCGCCGG
  4     AJ536231  AACTGTTGCTTCGGCGGGTCGGGGGCGC----CCTCCCGGGGGCGCCCCCCATGCCCGGCCCCT---CCC--------GGGGCCGAACGCCCGCCGG
  5     AY376527  -ACTGTTGCTTCGGCGGGTCGGGGGAC----CCCGT---------GCCCCCCCCCCGGCCGCG---CCCC---TCGCGGGGCGTGGCGCCCGCCGG
  6     AY376539  AACCGTTGCTTCGGCGGGTTAGGGGGTCCCCTCTCCGGGGGACGCCCTCCCGGCCGGGCCCCA-----CTGCGGG-----GCTCGGCGCCCGCCGG
  7     AY376542  AACTGTTGCTTCGGCGGGTTAGGGGGTCCCCCC----GGGGACGCCCTCCCGGCCGCGCCCTCCTCCTCC--GGGAGGGGTCGCGGCGCCCGCCGG
  8     AB105959  AACTGTTGCTTCGGCGGGCAGGAGGACAACCCCCCTCGGGGGGCGGTCCCCCTCCCGGCCGCG---CCC----TCACGGG-CGTGGCGCCCGCCGG
  8     AJ301938  AACTGTTGCTTCGGCGGGCAGGAGGACAACCCCCTC--GGGGGCGGTCCCCCTCCCGGCCGCG---CCC----TCACGGG-CGTGGCGCCCGCCGG
  9     AY376528  -ACTGTTGCTTCGGCGGGCAGGGGGTG----CCGCCTGC-GGACCCCCCT---CCCGGCCCT----GCCCT----CACGGGCGGAGCGCCCGCCGG
  9     AJ301955  -ACTGTTGCTTCGGCGGGCAGAGGTCC----CTTCT----GGGGACCCCT---CCCGGCCCCG---TCCCT----CGCGGACGGAGCGCCCGCCGG
 10     AY376510  AACCGTTGCTTCGGCGGGCAGGGGAAG----CCTCTCGCGGGCCTCCCCT---CCCGGCGCCG---GCCCCACCACGGGGACGGGGCGCCCGCCGG
 11     AJ301941  AACCGTTGCTTCGGCGGGC--GGGAGG----TCC-------GCCTCCCCC---CC-GGCCCCG---CTC----------GCGGGGCGCCCGCCGG
 11     AJ301958  AACCGTTGCTTCGGCGGGC--GGGAGG----TCC-------GCCTCCCCC---CT-G-CCCCG---CTC----------GCGGGGCGCCCGCCGG
 11     AY376541  AACCGTTGCTTCGGCGGGC--GGGAGG----TCC-------GCCTCCCCC---CCCGGCCCCG---CTC----------GCGGGGCGCCCGCCGG
 12     AY438547  AACTGTTGCTTCGGCGGGT-AGGCCGTCCCCTGA--AAAGGACG-CCTCCCGGCCCGGACCGGACCCHCCGCGGGACCGGACCCGGCGCCCGCCGG
 13     AY376531  AACTGTTGCTTCGGCGGGT-AGGC-GTCCCCTAA--AAAGGACG-TCTCCCGGCCCTCTCCCG----TCCGCGGG--TGG----GGCGCCCGCCGG
                   ** ***** ********                          *                                  **********
                            44
```

Fig. 2 Example of variable regions associated with *Colletotrichum* ITS1 structure types. Flanking inverted repeat sequences boxed

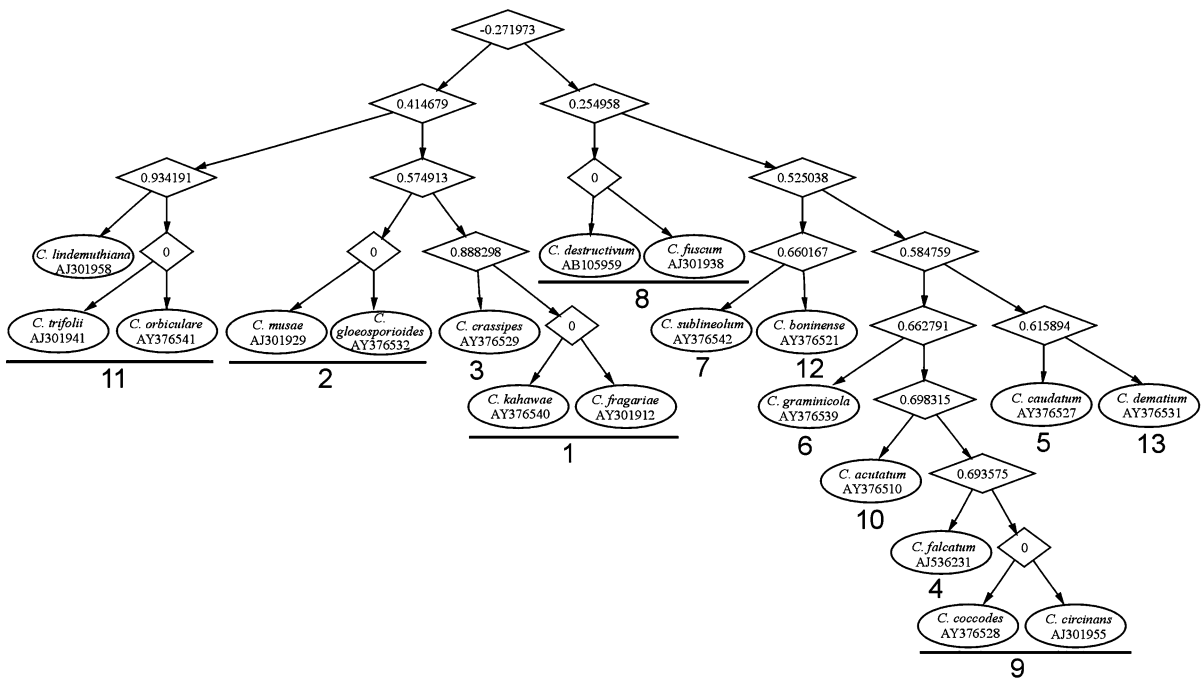


Fig. 3 RNAForester alignment of *Colletotrichum* ITS1 hairpins. Structure types 1–13 as indicated as in Fig. 1 and similarity scores given at nodes (note that identical structures and sequences do not give a similarity value and negative scores generally indicate that deletions may be needed for alignment)

types 1 and 2 in the *C. gloeosporioides* complex were recovered as 2 structure type subclusters with the exception of the sequences from *C. musae* and *C. crassipes*. These sequences had a type 2 and 3 structures, but grouped with the type 1 sequences. A second cluster contained the remaining non *C. gloeosporioides* species.

Trichophyton Sequences

All predicted models of the ITS1 structures for most of the *Trichophyton* and *Arthroderma* sequences showed a conserved double hairpin. There were four distinct structure types, each differing in the number and positions of loops in the double stem structure

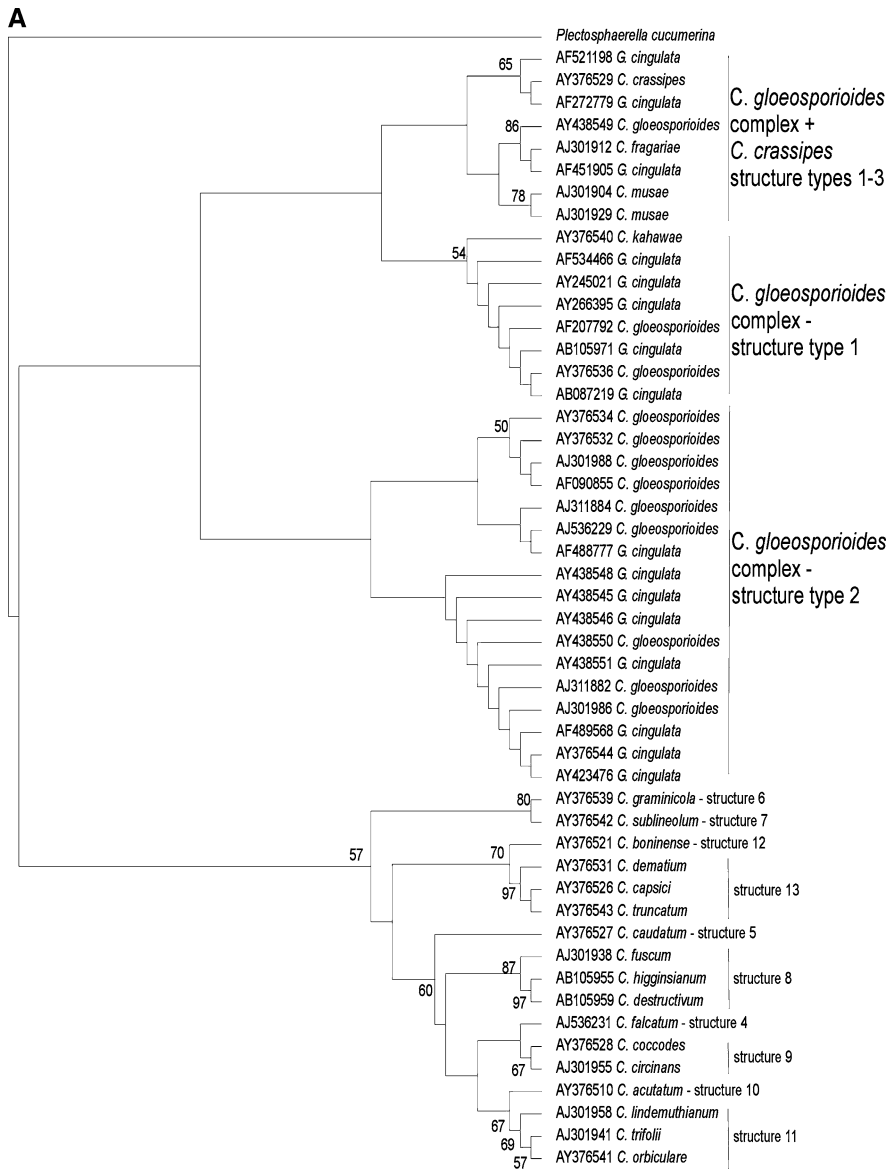
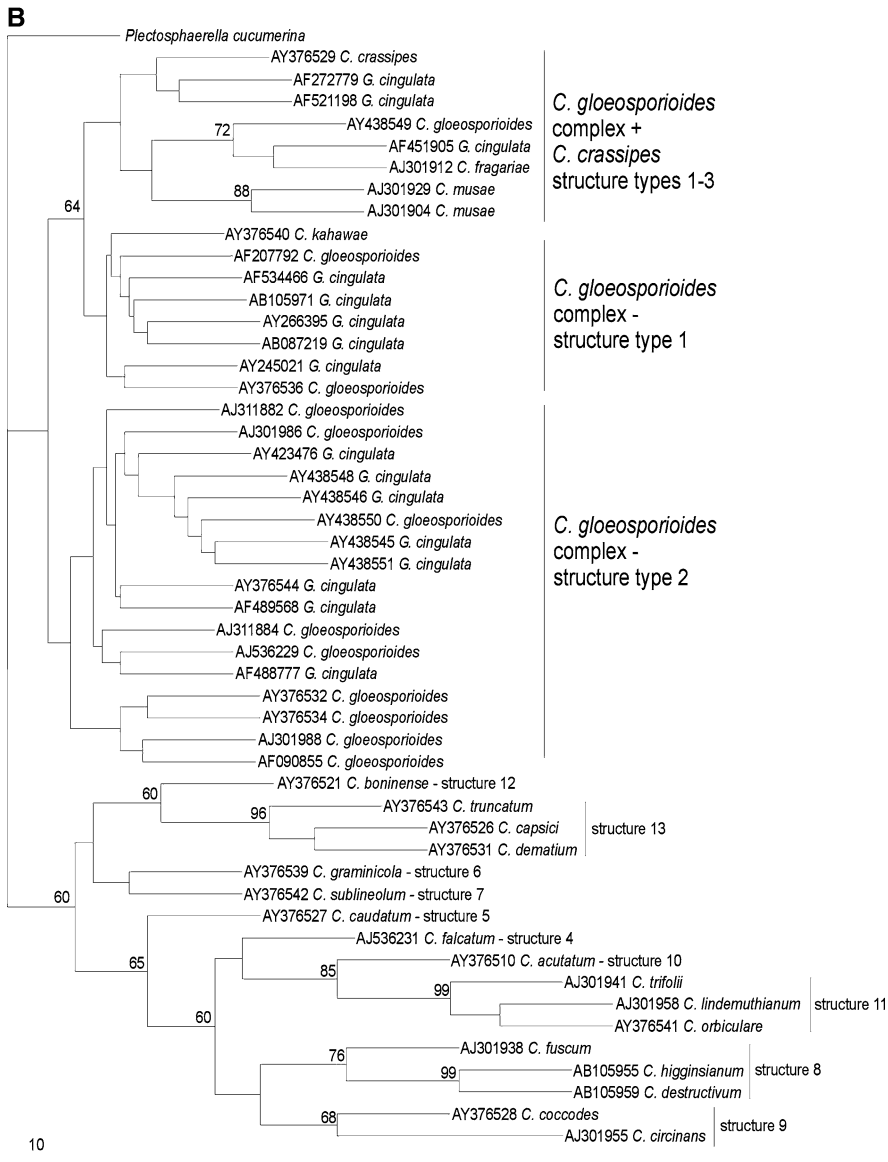


Fig. 4 (**a**) Consensus Neighbour Joining tree of complete ITS1 sequences from *Colletotrichum* species (bootstrap values above 50% shown). ITS1 hairpin structure type shown after species name. (**b**) Consensus Maximum Likelihood tree of complete ITS1 sequences from *Colletotrichum* species (bootstrap values above 50% shown). ITS1 hairpin structure type shown after species name

(Fig. 5) The sequences from this region were aligned separately and conserved and variable regions were identified. The variable sequence within the hairpin structure was not entirely conserved for each structural variant, and the four main structural variants contained eight sequence variants (Fig. 6). The hairpin structure consisted as in *Colletotrichum*, of a very variable GC rich sequence (70–82%G+C) of simple repeats and single base regions flanked by the CGGCGG/CCGCCG inverted repeat at positions 111 and 178 (with reference to Z98002). The exceptions to this were sequences labelled as from *T. terrestre*. Two of these (NCPF362 & 467) showed two different conserved double hairpins, similar to those from the other species, but the hairpin in the other two isolates (NCPF602 & 469) was not completely conserved, with only the region between bases 135 and 171 being conserved as a single stem (see Fig. 4). The sequence from NCPF362 contained a single transition (A for G) at position 6 in the first repeat, without a compensatory mutation in the inverted copy. The other three *T. terrestre* sequences contained a variable region in the same position as the

Fig. 4 continued



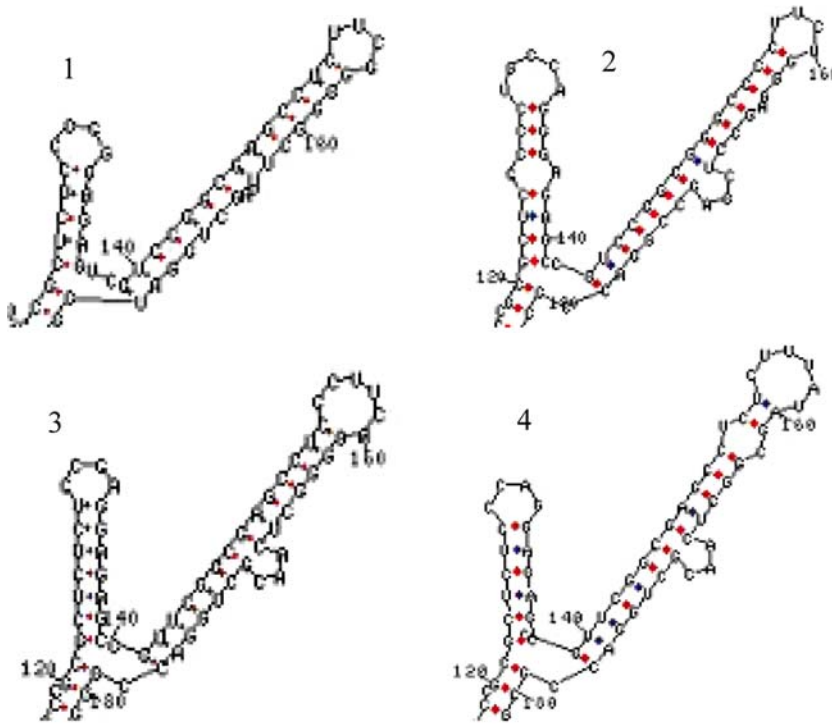
other species with a T in position 6 of the first repeat, and a matching A in position 1 of the inverted repeat (CGGCGT/ACGCCG).

Alignment of the different *Trichophyton* sequence/structure types through RNAForester gave a basal bifurcation in the *T terrestre* structures and recovered the other structures in a single line consisting of four subgroups (Fig. 7). Structure types 2 and 4 were recovered separately whereas the single type 3 structure was grouped with type 1.

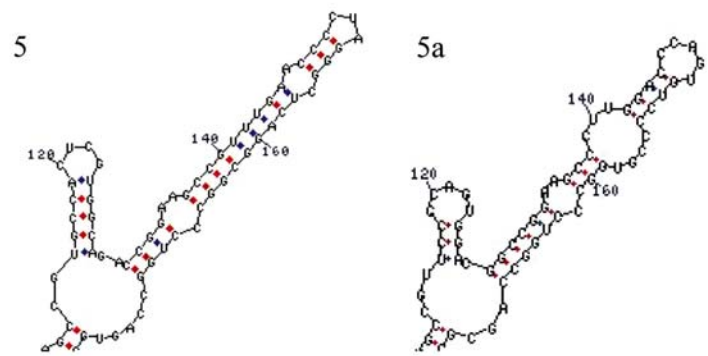
ITS2 sequences were not available for all of the *T. terrestre* strains. However, a neighbour joining tree of the complete ITS1 region for these isolates gave three major groups (Fig. 8). One comprised the sequences from *T. terrestre*, and the others comprised subgroups that corresponded with the original species identifications. Bootstrap values were uniformly high for all of the subdivisions.

Outgroups

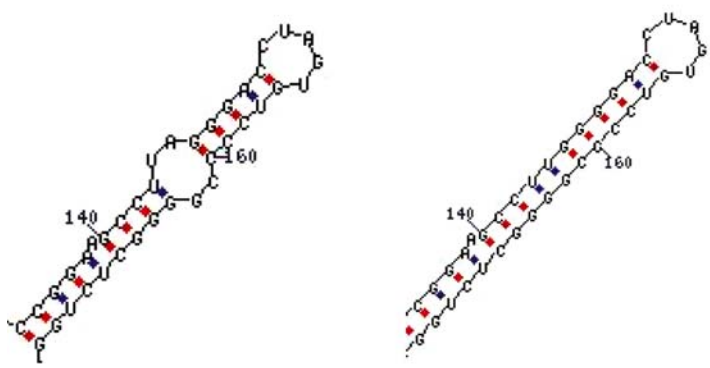
The common CGGCGG/CCGCCG inverted repeats were also effectively present in the ITS1 regions of the two sequences used as outgroups (*Plectosphaerella cucmerina* and *Microsporum canis*). In each



a. Four main structure groups obtained for anthropohilic *Trichophyton* species



b. Conserved double hairpin structures in geophilic *T. terrestre* NCPF 362 & 467



c. Conserved single stems in geophilic *T. terrestre* NCPF 469 & 602

Fig. 5 Detail of ITS1 structure types in *Trichophyton*

```
Structure
type    Sequence
  1     Z98002          TCGGCGGGCCGCGCT-CTCCC-----CGGAGAGTCGTCCGGCGAGCCTCTTC----GGGGGCTTTA-GCTGGATCGCGCCCGCCGG
  1     AB088676        TCGGCGGGCCGCGCT-CTCCT-----GGGAGAGTCGTCCGGCGAGCCTCTTT----GGGGGCTTTA-GCTGGATCGCGCCCGCCGG
  1     AB088678        TCGGCGGGCCGCGCT-CTCCT-----GGGAGAGTCGTCCGGCGAGCCTCTTC----GGGGGCTTTA-GCTGGATCGCGCCCGCCGG
  2     AJ270808        TCGGCGGGCCGCGCT-CCCCCTGCCAGGGAGAGCCGTCCGGCGGGCCCCTTCTG--GGAGCCTCGA-GCCGGACCGCGCCCGCCGG
  2     AF170464        TCGGCGGGCCGCGCT-CCCCCTGCCAGAGAGAGCCGTCCGGCGGGCCTCTTCCG--GGGGCCTCGA-GCCGGACCGCGCCCGCCGG
  3     AF170454        TCGGCGGGCCGCGCT-CTCTC---CCAGGAGAGCCGTTCGGCGAGCCTCCCTTC--AGGGGCTCAACGCTGGACCGCGCCCGCCGG
  4     Z98008          TCGGCGGGCCGCGCT-CTC-----CCAGGAGAGCCGTTCGGCGAGCCTCTCTTTATAGCGGCTCAACGCTGGACCGCGCCCGCCGG
  4     Z98001          TCGGCGGGCCGCGCT-CTC-----CCAGGAGAGCCGTTCGGCGAGCCTCTCTTT--AGTGGCTAAACGCTGGACCGCGCCCGCCGG
  5     362T_TERRESTRE  TCGGCGAGCCGTGCCACTCGTGGCAGACCGGAAGCCGTTTGAACCCCTAGGGCTCAGGCGGCC-----CTGGCCAGTGCCCGCCGA
  5a    467T_TERRESTRE  TCGGCGTGCCGTTCC-CCAGTGGACGGCCGGAAGCCCTTGGA---CCCAGTGTCCCCGTGGCC-----CTGGCCAGCGCACGCCGA
Single  469T_TERRESTRE  TCGGCGTGCCGTTCC-CCAGTGGACGACCGGAAGCCTTAGGGA--CCTAGTGTCCCCGGGGCT-----CTGGCCAGCGCACGCCGG
Single  602T_TERRESTRE  TCGGCGTGCCGTTCC-CCAGTGGACGACCGGAAGCCTTGGGGA--CCTAGTGTCCCCGGGGCT-----CTGGCCAGCGCACGCCGG
                        *****_**** **  *  *            **  *         **           * *      * **   * ** *****
                        111
```

Fig. 6 Example of variable regions associated with *Trichophyton* ITS1 structure types. Flanking inverted repeats sequences boxed

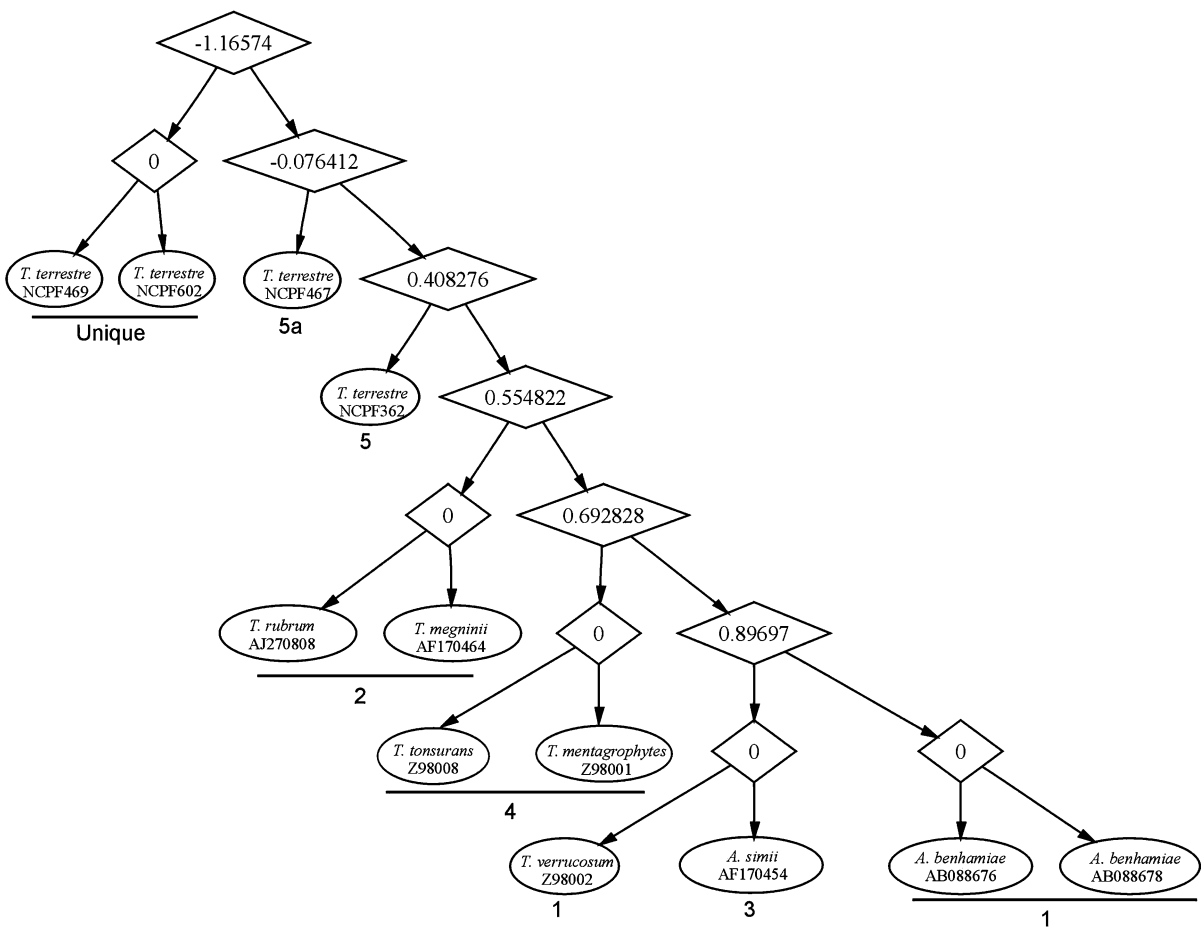


Fig. 7 RNAforester alignment of *Trichophyton* ITS1 hairpins. Structure types 1–5 indicated as in Fig. 5 and similarity scores given at nodes (note that identical structures and sequences do not give a similarity value and negative scores generally indicate that deletions may be needed for alignment)

case, there was a single difference with *P. cucumerina* having a G/C transversion at position 1 of the second repeat, and *M. canis* having a C/T transition in position 2 of the second repeat. In both cases, these repeats formed the base of loops of variable sequences equivalent to those obtained from the

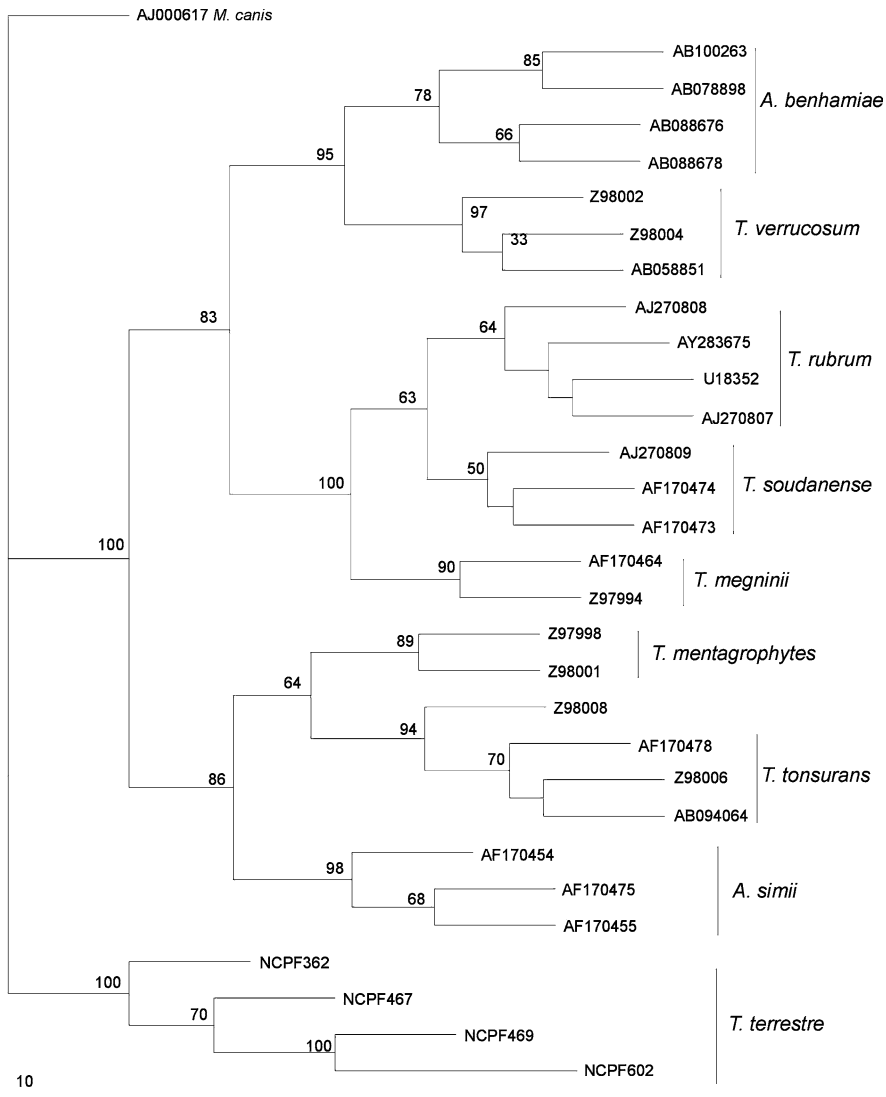


Fig. 8 Consensus Neighbour Joining tree of ITS1 sequences from *Trichophyton* species (bootstrap values above 50% shown). ITS1 hairpin structure type shown after species name

Colletotrichum and *Arthroderma* species (structures not shown).

Discussion

There is generally a good correlation between groupings obtained from the structure analysis and the groupings obtained from the sequence analysis. As the sequences giving rise to the structures are sub-sets of those used in sequence analysis, this could indicate that much of the current sequence based analysis could be traced back to short specific regions of the ITS. In both taxa, some species showed unique structures and some species showed the same structure. In nearly all cases, the same differences and similarities were also obtained through sequence analysis and also reflected current opinions on the taxonomy of the species.

Colletotrichum

Sreenivasaprasad et al. [7] identified most variation in the ITS1 regions of *Colletotrichum* between 60 and 120 bp positions. This is the region that includes the hairpin, and so the structural groupings

could be expected to support ITS1 sequence placements. This is the case for *C. gloeosporioides, C. musae* and *C. kahawae* (structure types 1 & 2), *C. orbiculare, trifolii* and *lindemuthiana* (structure type 11), *C. graminicola* and *C. sublineolum* (structure types 6 & 7), and *C. truncatum* and *C. dematium* (structure type 13) that were grouped together by Sreenivasaprasad et al. [7] and Freeman et al. [29], although the latter grouping was not supported by ITS2 analysis [30]. However *C. fuscum* and *C. capsici*, species that had been grouped with *C. gloeosporioides* by ITS1 sequences [7, 29], each have different hairpin structures (types 8 & 13). An interesting finding is the division of *C. gloeosporioides* sequences into 2 structure groups. Each structure type consisted of two slightly different variable region sequences, but there was no obvious consistent difference between the two structure types. However, both structure types were recovered as single structure subgroups in the NJ tree with the exception of *C. musae* and *C. crassipes*. Sreenivasaprasad et al. [7] reported low bootstrap values and up to 4% ITS1 sequence difference within *C. gloeosporioides*, and a similar range was found here for a larger number of sequences. It is not possible to directly infer phylogeny from the alignment methods available for comparing the structures, but the consistent difference between structure types 1 and 2 may indicate some more cryptic separation.

One concern from the results obtained here are the relatively low levels of bootstrap support seen in the *Colletotrichum* tree. In general terms, only the branch points grouping species or small groups of species were statistically supported. Similar levels of support have been obtained in other studies with *Colletotrichum* (e.g. [19]) and this may account for the lack of consistency in the placement of species groups between studies. This suggests, although the ITS sequences can be used reliably to support groupings at the level of species or species complex, the spread of, and level of, variation within them is insufficient to determine relationships within or between these groups. This has been highlighted previously, and it has been shown that in some cases, nucleotide sequence distances within species groups may be greater than those between species [19]. Further analysis of the sequence data with maximum likelihood methods (Fig. 4) recovered the sequences in the same groups of 1–3 species, but with apparently different relationships between some of these, and this supports the unreliability of ITS sequences in determining phylogenetic relationships between species groups and complexes in this genus.

Trichophyton

The groupings recovered in the ITS1 tree (Fig. 8) and the hairpin structure type (Figure 7) correspond well with the currently accepted taxonomy. The linking of *A. benhamiae* with *T. verrucosum* (structure type 1), *T. rubrum* with *T. soudanense* and *T. megninii* (structure type 2) and *T. mentagrophytes* with *T. tonsurans* (structure type 4) are in agreement with the previous studies from ITS1 by Summerbell et al. [20] and Makimura et al. [8, 31]. Both of these authors placed *A. simii* in the general *T. mentagrophytes* species complex, although Summerbell et al. placed the species as a sister lineage to *A. benhamiae/T. soudanense/T. rubrum* [20], and Makimura et al. placed the species close to *T. tonsurans* and *T. mentagrophytes* [8]. This latter arrangement was supported by Gräser et al. [32] with complete ITS1/5.8S/ITS2 sequences who also placed the species in the same clade as *T. tonsurans* and *T. mentagrophytes*. This latter view is supported by the ITS1 tree here, but the hairpin tree placed *A. simii* (structure type 3) closer to *T. verrucosum.* One explanation is that the structure models used here are based on similarities across a single structural feature, and do not include any small differences in other parts of the molecule. The ITS 1 sequences of *A. benhamiae* and *T. verrucosum* differ by only two base pairs and so these species would have been expected to group together, separate from *A. simii* which has a longer variable region. As the variable region is within the hairpin structure, it will have a greater effect on comparisons. Structures from sequences from the geophilic *T. terrestre* complex appeared as a loose group at the top (root) of the structure tree. This was mirrored in the ITS tree and may suggest a basal similarity to the other species. One possible explanation that is supported by the range of hairpin structures obtained for *T. terrestre* isolates is that the anthrophilic species have been derived from the *T. terrestre* complex.

Methodology

Previous analyses of ITS sequences for these fungi have usually been made from a variety of phylogenetic methods including parsimony and maximum likelihood (e.g. [7, 8]). The RNAForester methodology constructs a progressive similarity based alignment of structures with nearest neighbours being joined first. This does not include any phylogenetic model and so in order to compare "like with like" distance measures and neighbour joining were used here. Similarly a "standard" model was used for the maximum likelihood trees as a single variation rate is the closest available to the progressive similarity used for the structural comparisons. The groupings obtained with the similarity based methods used here and elsewhere appear to be comparable with groupings published from phylogenetic methods (e.g. [19, 32]).

The generation of secondary structural models is a predictive process, and so the structures obtained can only be considered as those that are most likely thermodynamically (e.g. [33]). There are dangers in assuming that the minimum free energy models are those produced in vivo, and so identification of consistent features and sequence analysis are important checks to these [10]. The consistent appearance of a structural feature in all of the most likely models, either intuitively or mathematically, has been taken as indicative of a conserved feature in various organisms [10, 15, 34]. In this study, freely available tools were used for the sequence and the structure comparisons. The identification of consistent structures and their initial relationship to the sequences was largely undertaken manually. This is a crucial part of the analysis, and although tools such as the JalView conservation function are available for the sequence comparisons, there does not appear to be an easily available utility suitable for identifying conserved structures between sequences and all probable models. It is therefore likely that until the analysis can be fully automated the use of such features will be limited to relatively small studies. Tuckwell et al. generated Hidden Markov Models for the subsequent identification of new sequences [15], but there may be constraints to this approach if longer variable regions are used.

There has been some discussion as to the accuracy of structures obtained from folding and alignment algorithms, and a number of methods have been compared recently [35]. Our approach used an initial ClustalW alignment step to derive homologous sequences from the ITS region, and to identify the common inverted repeat. This then allowed us to identify that there was a hairpin structure and to place this within the proposed ITS1 structure [10, 12, 15]. This approach allowed us to start our structural comparisons from MFold and RNAForester with known homologous regions and structures. This is most analagous to the 'plan C' described by Gardner and Giegerich [35]. The sequences of the variable regions used here were frequently very different from each other and the finding that RNAForester performed better with medium and low similarity data [35], also supports its use in this case.

Structure Identity

Short repeated sequences have been reported in numerous eukaryote ITS regions. Campbell et al. [12] described short subrepeats flanking highly conserved sequences in *Picea* and Won and Renner [13] have reported tandem repeats in the ITS of *Gnetum*. It has been suggested that ITS sequences flanked by such conserved repeats could be paralogues of pseudo-genes [36], and in the fungi described here the intervening sequences are very variable, but all have very high G/C ratios. The relationship of such motifs to the secondary structure has only been considered in a small number of instances. Gottschling and Plötner [10] identified conserved motifs in the loops associated with ITS1 helices in dinoflagellates, and Campbell et al. [12] found large hairpin structures associated with short subrepeats. In the study of the rumen gut associated chytrids, all sequences were from a single chytrid family and four variable hairpin structures anchored by inverted repeats were used for genus/species differentiation [15]. The inverted repeats found in the ascomycete fungi studied here conform to the GGCRY-RYGYC motif that has been identified as the stem of helix 1C in some flowering plants [34, 37], and a similar inverted repeat has also been described in domain II of the ITS1 in some ascomycetous yeasts [9]. In that study, the hairpin was described as "an evolutionarily highly conserved element" and it was suggested that it played an unknown role in ribosome biogenesis. Alterations to the size of the stem and the sequence of

the loop do not apparently affect the production of the mature rRNA [9] and so this could provide a constrained region where sequence variation could accumulate. In plants and yeasts the inverted repeats were only separated by short variable sequences, and not the longer variable regions obtained here. The inverted repeats and similar variable regions studied in *Colletotrichum* and *Trichophyton* were also detected in the fungi used as outgroups. This could indicate that they are found outside the taxa used here and that they may have wider applications in mycology.

Conclusion

This study has demonstrated that species groupings made by both ITS1 and complete ITS region sequences are supported by the structure of the ITS1 domain II conserved hairpin motif. Structural analyses grouped related species, and in *C. gloeosporiodes* gave greater resolution than conventional sequence analysis. There was also some suggestion of a correlation between structural type and morphology. In *Trichophyton* structural analysis produced groups that correlated with proposed phylogenetic lines. These correlations suggest that the structure of the domain II hairpin provides an additional tool for interpreting high levels of ITS sequence homology between closely related fungi.

References

1. Bridge PD, Spooner BM, Roberts PJ. The impact of molecular data in fungal systematics. Adv Bot Res 2005;42:33–67.
2. Edel V, Steinberg C, Gautheron N, Alabouvette C. Evaluation of restriction analysis of polymerase chain reaction (PCR)-amplified ribosomal DNA for the identification of *Fusarium* species. Mycol Res 1996;101:179–87.
3. Bridge PD, Waller JM, Davies D, Buddie AG. Variability of *Colletotrichum kahawae* in relation to other *Colletotrichum* species from tropical perennial crops and the development of diagnostic techniques. J Phytopathol 2008. doi:10.1111/j.1439-0434.2007.01354.x.
4. Tymon AM, Shah PA, Pell JK. PCR-based molecular discrimination of *Pandora neoaphidis* isolates from related entomopathogenic fungi and development of species-specific diagnostic primers. Mycol Res 2004;108:419–33.
5. Xue B, Goodwin PH, Annis SL. Pathotype identification of *Leptosphaeria maculans* with PCR and oligonucleotide primers from ribosomal internal transcribed spacer sequences. Physiol Mol Plant Pathol 1992;41:179–88.
6. Brown AE, Muthumeenakshi S, Sreenivasaprasad S, Mills PR, Swinburne TR. A PCR primer-specific to *Cylindrocarpon heteronema* for detection of the pathogen in apple wood. FEMS Microbiol Lett 1993;108:117–20.
7. Sreenivasaprasad S, Mills PR, Meehan BM, Brown AE. Phylogeny and systematics of 18 *Colletotrichum* species based on ribosomal DNA spacer sequences. Genome 1996;39:499–512.
8. Makimura K, Tamura Y, Mochizuki T, Hasegawa A, Tajiri Y, Hanazawa R, Uchida K, Saito H, Yamaguchi H. Phylogenetic classification and species identification of dermatophyte strains based on DNA sequences of nuclear ribosomal internal transcribed spacer 1 regions. J Clin Microbiol 1999;37:920–4.
9. van Neus RW, Rientjes JM, van der Sande CA, Zerp SF, Sluiter C, Venema J, Planta RJ, Raue HA. Separate structural elements within the internal transcribed spacer 1 of *Saccharomyces cerevisiae* precursor ribosomal RNA direct the formation of 17S and 26S rRNA. Nucl Acid Res 1994;22:912–9.
10. Gottschling M, Plötner J. Secondary structure models of the nuclear internal transcribed spacer regions and 5.8S rRNA in Calciodinelloideae (Peridiniaceae) and other dinoflagellates. Nucl Acid Res 2004;32:307–15.
11. Lalev AI, Nazar RN. Structural equivalence in the transcribed spacers of pre-rRNA transcripts in *Schizosaccharomyces pombe*. Nucl Acid Res 1999;27:3071–8.
12. Campbell CS, Wright WA, Cox M, Vining TF, Major CS, Arsenault MP. Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in Picea (Pinaceae): sequence divergence and structure. Mol Phylogen Evol 2005;35:165–85.
13. Won H, Renner SS. The internal transcribed spacer on nuclear ribosomal DNA in the gymnosprem *Gnetum*. Mol Phylogenet Evol 2005;36:581–97.
14. Gonzalez P, Labarère J. Sequence and secondary structure of the mitochondrial small-subunit rRNA V4, V6 and V9 domains reveal highly species-specific variations within the genus *Agrocybe*. Appl Environ Microbiol 1998;64:4149–60.
15. Tuckwell DS, Nicholson MJ, McSweeney SS, Theodorou MK, Brookman JL. The rapid assignment of ruminal fungi to presumptive genera using ITS1 and ITS2 RNA secondary structures to produce group-specific fingerprints. Microbiology 2005;151:1557–67.
16. Landis FC, Gargas A. Using ITS2 secondary structure to create species-specific oligonucleotide probes for fungi. Mycologia 2007;99:681–92.
17. Cannon PF, Bridge PD, Monte E. Linking the past, present, and future of *Colletotrichum* systematics. In: Prusky D, Freeman S, Dickman M, editors. *Colletotrichum*: host specificity, pathology, and host-pathogen interaction. St Paul Minnesota: American Phytopathological Society, 2000. p. 1–20.
18. Sreenivasaprasad S, Brown AE, Mills PR. DNA sequence variation and interrelationships among *Colletotrichum* species causing strawberry anthracnose. Physiol Mol Plant Pathol 1992;41:265–81.
19. Sreenivasaprasad S, Mills PR, Brown AE. Nucleotide sequence of the rDNA spacer 1 enables identification of

isolates of *Colletotrichum* as *C. acutatum*. Mycol Res 1994;98:186–8.
20. Summerbell RC, Haugland RA, Li A, Gupta AK. rRNA gene internal transcribed spacer 1 and 2 sequences of asexual, anthropophilic dermatophytes related to *Trichophyton rubrum*. J Clin Microbiol 1999;37:4005–11.
21. Kirk PM, Cannon PF, David JC, Stalpers JA, editors. Ainsworth & Bisby's dictionary of the fungi. 9th ed. Wallingford: CAB International; 2001.
22. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acid Res 1994;22:4673–80.
23. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucl Acid Res 2003;31:3406–15.
24. Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java Alignment Editor. Bioinformatics 2004;20:426–7.
25. Höchsmann, M. The tree alignment model: Algorithms, implementations and applications for the analysis of RNA secondary structures. PhD Thesis Bielefeld University, 2005. (available as .pdf file at http://bibiserv.techfak.uni-bielefeld.de/rnaforester/diss_hoechsmann.pdf).
26. Höchsmann M, Voss B, Giegerich R. Pure multiple RNA secondary structure alignments: a progressive profile approach. IEEE/ACM Trans Comput Biol Bioinform 2004;1:53–62.
27. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, 2005.
28. Page RDM. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Appl Biosci 1996;12:357–8.
29. Freeman S, Minz D, Jurkevitch E, Maymon M, Shabi E. Molecular analyses of *Colletotrichum* species from almond and other fruits. Phytopathology 2000;90:608–14.
30. Afanador-Kafuri L, Minz D, Maymon M, Freeman S. Characterization of *Colletotrichum* isolates from Tamarillo, Passiflora, and Mango in Colombia and identification of a unique species from the genus. Phytopathology 2003;93:579–87.
31. Makimura K, Mochizuki T, Hasegawa A, Uchida K, Saito H, Yamaguchi H. Phylogenetic classification of *Trichophyton mentagrophytes* complex strains based on DNA spacers of nuclear ribosomal internal transcribed spacer 1 regions. J Clin Microbiol 1998;36:2629–33.
32. Gräser Y, de Hoog GS, Kuijpers AFA. Recent advances in the molecular taxonomy of dermatophytes In: Kushwaha RKS, Guarro J, editors. Biology of Dermatophytes and other Keratinophilic fungi. Bilbao Spain: Revista Iberoamericana de Micologia (suppl.), 2000, p. 17–21.
33. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. JMol Biol 1999;288:911–40.
34. Goertzen LR, Cannone JJ, Gutell RR, Jansen RK. ITS secondary structure derived from comparative analysis: implications for sequence alignment, phylogeny of the Asteraceae. Mol Phylogenet Evol 2003;29:216–34.
35. Gardner PP, Giegerich R. A comprehensive comparison of comparative RNA structure prediction approaches. BMC Bioinformatics 2004;5:140.
36. Mayol M, Rossello JA. Why ribosomal DNA spacers (ITS) tell different stories in Quercus. Mol Phylogenet, Evol 2001;19:167–76.
37. Liu JS, Schardl CL. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Mol Biol 1994;26:775–8.