Sampling and sequencing

Sampling and morphometric analysis

Fungal specimens were collected throughout Thailand. The date and location of all collections, including type of host, were recorded. Fresh stromata were photographed and described, and if appropriate, spores were removed and transferred onto Potato Dextrose Agar (PDA) for culturing. The morphological characteristics of the spores such as shape, size, and color of stromata, perithecia, ascospores, etc. were observed [\[1-4\]](#page-13-0). The color reaction between 10% Potassium hydroxide (KOH) and stromatal surface was performed and compared to the mycological color chart [\[5\]](#page-13-1). Ascospores mounted in Melzer's reagent were examined for characteristics such as the type of germ slit and then photographed by light microscopy (Olympus BX-40). Perispores and epispores in 10% KOH were observed. The size range of the spores was calculated from measurements of the width and length of at least thirty spores. Any ornamentation present was recorded and later critically analyzed using scanning electron microscopy (SEM, JEOL-840). All fungal collections are located at the Microbiology Programme Herbarium, Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand.

PCR and sequencing

Fungal mycelium was harvested from pure cultures on PDA. Mycelium was washed twice with sterile distilled water, frozen at -70°C and ground with a micro pestle. Fungal genomic DNA was extracted following the instructions of a DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA). The ITS nrDNA was amplified by using ITS5/ITS4 as forward and reverse primers, respectively (66). PCR reactions were carried out in a total volume of 50 µl containing 100 ng DNA template, 1 X PCR buffer and 1.5 mM $MgCl_2$, 2 mM dNTPs, 0.2 µM of each primer and 0.5 units of Taq Polymerase (Invitrogen, USA). Thermal cycling parameters were as following; initial denaturation at 95 °C for 5 mins, followed by 30 cycles of 95 °C for 30 sec, 55 °C for 30 sec, 72 °C for 1 min and a final elongation for 7-10 mins at 72 °C. Further problematic isolates were also studied on the partial region of α-actin and β-tubulin genes using ACT512F/ACT783R (67) and T1/T22 [\[6\]](#page-13-2) as primers, respectively. The parameters of thermal cycling were as following; initial denaturation at 95° C for 5 mins, followed by 30 cycles of 95° C for 1 min, 50 $^{\circ}$ C for 1 5 min 72 $^{\circ}$ C for 2 min and a final alongstion for 10 mins at 72 $^{\circ}$ C. The PCP products C for 1.5 min, 72 $\rm{^oC}$ for 2 min and a final elongation for 10 mins at 72 $\rm{^oC}$. The PCR products were checked in 1 % agarose gel electrophoresis before purified with QIAquick PCR Purification Kit (QIAGEN). They were then directly sequenced with BigDye on an ABI 377 automated sequencer (Perkin Elmer, USA) using the same primers as for amplification. Each sequence was manually checked for ambiguous bases and the two sequences (direct and reverse) were assembled with Bioedit 5.0.6 (69).

Summary statistics for the ITS sequence data

The original dataset consisted of 269 FASTA records. As described at the start of the Results section, indeterminate identification required exclusion of four *Annulohypoxylon* sequences from the barcode analyses, leaving our complete dataset with 265 FASTA records and 213 unique sequences.

Within the complete dataset, there were 94 samples from 16 of the 38 *Annulohypoxylon* species, 167 samples from 41 of the 112 *Hypoxylon* species, and 4 sequences under *Nemania serpens*. Of the 16 *Annulohypoxylon* species, 12 had at least 2 samples; of the 41 *Hypoxylon* species, 23. Sequence lengths ranged from 384 to 936, with the mean \pm standard error equaling 556 \pm 116.

To detect biases in sequences from GenBank, the complete dataset was subdivided into GenBank and non-GenBank datasets.

The GenBank dataset consisted of 128 FASTA records with 103 unique sequences. There were 42 samples from 11 *Annulohypoxylon* species, 82 samples from 25 *Hypoxylon* species and the 4 *Nemania serpens* samples. Of the 11 *Annulohypoxylon* species, 6 had at least 2 samples; of the *Hypoxylon* species, 8 had at least 2 samples. Sequence lengths ranged from 384 to 936, with the mean \pm standard error equaling 547 \pm 96.

The non-GenBank dataset contained 137 FASTA records with 110 unique sequences. There were 52 samples from 10 *Annulohypoxylon* species and 85 samples from 24 *Hypoxylon* species. Of the 10 *Annulohypoxylon* species, 8 had at least 2 samples; of the *Hypoxylon* species, 19 had at least 2 samples. Sequence lengths ranged from 444 to 916, with the mean ± standard error equaling 567 ± 132 .

In both GenBank and non-GenBank datasets, each unambiguous nucleotide (A,C,G,T) constituted an unremarkable 22% to 28% of the unique sequences; all ambiguous nucleotides together, less than 0.1%.

Supplementary phylogenetic analyses

The phylogenetic tree recovered many *Annulohypoxylon* and *Hypoxylon* species in highly supported monophyletic clades, supporting the use of ITS nrDNA sequences as barcodes for these conflictive groups of fungi. *N. serpens* proved best of several outgroups tried for the tree, although the tree placement of the *Annulohypoxylon* and *Hypoxylon* genera included the *N. serpens* sample AY805565 (from GenBank, and therefore lacking a standard voucher). Most sequences in the blue clade contained only one "species name" such as *A. annulatum*, *A. atroroseum*, *A. bovei* var. *microspora*, *A. cohaerens*, *A. multiforme*, *A. nitens*, *A. purpureonitens*, *A. stygium*, *A. truncatum*, *H. diatrypeoides*, *H. duranii*, *H. fendleri*, *H. fragiforme*, *H. fuscum*, *H. haematrostroma*, *H. investiens*, *H. lenormandii*, *H. macrocarpum*, *H. monticulosum*, and *H. sublenormandii*. Fourteen sequences of *A. annulatum* were grouped together and shared the same host of *Tremella fuciformis* noted from China except that two sequences of *A. annulatum* (GB114; AJ390395 and GB105; AM749938) were segregated in another two different clades. One *A. annulatum* sequence (GB105; AM749938) exhibited the extremely long ITS1 region (584 bp) which is similar to those found in *A. atroroseum*, *A. stygium* and *"A.* aff. *nitens"* specimens, whereas another *A. annulatum* sequence (GB114; AJ390395) showed the common size (179 bp). There were three *Annulohypoxylon truncatum* sequences grouped together

including one *A. annulatum* sequence (GB114; AJ390395), which had identical sequences to *A. truncatum* of up to 99% similarity. In the *Annulohypoxylon cohaerens* clade one sequence of *A. minutellum* (GB124: AJ390399) was included with three sequences of *A. cohaerens*. This result indicates the close relationship within this group since *A. minutellum* has been published as a synonym of *A*. *cohaerens* var. *microsporum* [\[7\]](#page-13-3). It has 87% similarity to all *A. cohaerens* sequences.

This phylogenetic analysis also resolved closely related species such as *A. atroroseum* and *A. stygium*, which are basically separated from each other by stromatal surface color and anamorph characters [\[2\]](#page-13-4). They were clearly separated by ITS nrDNA sequences. Similarly, *A. bovei* var. *microspora*, *A. moriforme* and *A. nitens* are quite close in morphological characters and presence of a *Nodulisporium*-like anamorph, but their ITS sequences distinguished them. The phylogenetic tree in Figure S2 also clearly separates the "cryptic species" *A. nitens* and "*A.* aff. *nitens*" (green clades). Under global alignment within an MSA, each of *A. nitens* and "*A.* aff. *nitens*" was correctly identified monophyletically, so as a barcode marker, ITS separates the cryptic species found in this study. Some species that possess distinct and generally recognized characters such as *A. purpureonitens*, *H. haematrostroma*, *H. lenormandii*, *H. sublenormandii* and *H. monticulosum* exhibited clear species separation in each of the clades. Particularly, *H. fragiforme* the type species of the genus *Hypoxylon* was commonly studied and had the highest numbers of ITS nrDNA sequences available in GenBank as shown in a large blue clade (29 sequences). However, another eight sequences of *H. fragiforme* were placed in 3 different clades. Four sequences (GB129; AJ246218, GB130; AJ246217, GB131; AJ246216 and GB132; AJ246215) grouped in a complex clade, which contained another 4 different species. Two *H. fragiforme* sequences (GB44; EU715619 and GB45; EU715613) were included in the *H. investiens* group, whereas the last two *H. fragiforme* sequences (GB86; EF155510 and GB69; EF1555527) were individually separated. In the case of *H. fendleri*, although the specimens studied revealed high variation in minor morphological characters as described above, they still grouped together in both the GenBank and non-GenBank datasets.

In the *A. archeri* clade most members belonged to the non-GenBank dataset, except for one sequences of *A. moriforme* var. *microdiscus* (GB16; EF026137). Both species are very close in morphological characters except that perithecial diameter of *A. archeri* (0.1-0.3 mm) is smaller than that of *A. moriforme* var. *microdiscus* (> 0.3 mm) [\[2\]](#page-13-4). They do however lack anamorph character information. Also a recent study of *alpha*-actin and *beta*-tubulin genes within *Hypoxylon* and closely related genera recognized only *A. moriforme* var. *microdiscus* [\[7\]](#page-13-3). Both species need more collections and extensive study of both teleomorph and anamorph characters to confirm or reject this conclusion.

Several clades of sequences containing species complexes are placed in orange clades. The largest of these groups included the five members of *H. perforatum* (GB107; AM749935, GB61; FJ464593, GB117; AJ390407, QD322118 and QD322120), *H. carneum* (GB110; AM749926), *H. shearii* var. *minor* (GB11; EF026142), *H. subdisciforme* (GB8; FJ172269: unpublished species) and *H.* cf. *pilgerianum* (H064). Although the *H. perforatum* sequences from both GenBank and non-GenBank datasets placed in this clade it also included four other species. The sequence of *H. carneum* (GB110; AM749926) was 99% similar to *H. perforatum* (GB107; AM749935), and both of them had the same chemotype (63). However, ITS sequences did not

support the species identification within this complex clade: *H. fragiforme* (GB129; AJ246218, GB130; AJ246217, GB131; AJ246216 and GB132; AJ246215), *A. cohaerens* (GB94; AY909025) and *A. multiforme* (GB20; AY616708 and GB21; AY616707) had similar sequences, exhibiting 98% sequence identity. In contrast, in another complex clade, *H. crocopeplum* (H119), *H. rickii* (GB116; AJ390408) and *H. cinnabarinum* (GB125; AJ390398) were different in many characters, and their ITS sequences showed up to 12% divergence. In yet another complex clade, consisting of *H. rubiginosum* (H080) and *H. subgilvum* var. *microsporum* (SUT108QD322122), there were two different sequences from two different specimens, indicating a need to examine more specimens.

H. investiens | GB4 | FJ185308 | - | Strain CBS 118185; New Zealand *H. investiens* **GB5 GB5** FJ185307 **FULLER FULLER FULLER GGB5 EVALUATION CGS** 118183; Malaysia *H. kanchanapisekii* SUT066 - QD223741 Ratchaburi Province, Thailand *H. kanchanapisekii* SUT068 - QD223742 Ratchaburi Province, Thailand *H. kanchanapisekii* | SUT069 | - | QD223743 | Ratchaburi Province, Thailand *H. lenormandii* $H063$ - FM209435 Phitsanulok Province, Thailand *H. lenormandii* | H212 | - | FM209463 | Phitsanulok Province, Thailand

Table S2: The NCBI BLASTN Default Scoring System.

The accompanying gap penalty is $g(i) = 5 + 2i$ for a gap of length *i*.

Any aligned letter-pair containing an ambiguous nucleotide was given score 0.

Table S3: The UCSC BLASTZ Default Scoring System.

The accompanying gap penalty is $g(i) = 400 + 30i$ for a gap of length *i*.

Any aligned letter-pair containing an ambiguous nucleotide was given score 0.

Table S4: Probability of Correct Identification (PCI) by Species.

The PCI uses multiple alignment and p-distance. Note: in *Annulohypoxylon cohaerens*, ITS analysis correctly identified only 0.25 of the samples, so the species as a whole is not correctly identified.

References

- 1. Hayashi K, Hatsutori N, Donho M, Nakajima H (1996) Dihydrofolate reductase inhibitors containing mitorubrin compounds for protozoiasis. Japanese Patent 08217673.
- 2. Ju Y-M, Rogers JD (1996) A revision of the genus *Hypoxylon*. Mycologia Mem 20: 1-365.
- 3. Ju Y-M, Rogers JD (2002) The genus *Nemania* (Xylariaceae). Nova Hedwigia 74: 75-120.
- 4. Miller JH (1961) A monograph of the world species of Hypoxylon. Athens, U.S.A.: University of Georgia Press.
- 5. Rayner RW (1970) A mycological colour chart. Commonwealth Mycological Institute, Kew.
- 6. O'Donnell K, Cigelnik E (1997) Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. Mol Phylogenetics Evol 7: 103–116.
- 7. Hsieh HM, Ju YM, Rogers JD (2005) Molecular phylogeny of *Hypoxylon* and closely related genera. Mycologia 97: 844-865.