Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Natural Resources
and Agricultural Sciences

RDNA Barcodesystem

Fredrik Götesson

Uppsala
2012

DNA Barcodesystem

Fredrik Götesson

Supervisors:

Björn Lidahl

E-mail: Bjorn.Lindahl@slu.se

Department of Forest Mycology and Plant Pathology

Katarina Ihrmark

E-mail: Katarina.Ihrmark@slu.se

Department of Forest Mycology and Plant Pathology

Examiner:

Malin Elfstrand

E-mail: Malin.Elfstrand@slu.se

Department of Forest Mycology and Plant Pathology

Course: Bachelor in Biology

Course code: EX0689

Level of work: Basic

Credits: 15

Programme/Education: NK002 Biotechnology – Bachelor Programme

Place of publication: Uppsala, Department of Forest Mycology and Plant Pathology

Year of publication: 2012

Abstract

Frequently found ITS sequences that don't match any species in GeneBank were further classified with two different barcode regions. This study aims to answer for which phylogenetic placement the unknown species belongs to and if the species belong to a new order. The two barcode regions that were used were 18S small ribosomal subunit and the 28S large ribosomal subunit. The study will as well investigate which of the small and the large ribosomal subunit is appropriate extending ITS in a two-marker barcode system. Problem with the PCR amplifications made it hard to amplify many of the samples. The sequenced sample was able to copy the 28S large ribosomal subunit. The phylogenetic analysis with sequences from all classes showed that the sample belonged to the group of Pezizomycotina fungus. The parsimony phylogenetic tree showed that the sequence was lined along species from the order of Arthoniales with a bootstrap value on 98. The branch of Arthoniales is divided into subbranches with species from the order of Dothideomycetes with a low confident value. The H0 sample was collected from needle samplings. Arthoniales is an order of lichens species which is unlikely to be found from needle samples. Most likely the sample belongs to a Dothideomycetes species which is highly similar to Arthoniales species. The 18S small subunit has fewer hypervariable than 28S subunit and by that it is more suitable for higher taxonomic identification. 28S subunit combined with the ITS region is considered to be a good universal barcode.

Introduction

The total amount of fungal species in the world has been estimated to around 1.6 million species. Comparing to the 43,000 known characterized fungi species (2011) the number of all fungi species is estimated to around 611.000. Up to now the discovered species are just like the tip of an ice berg comparing to the total estimations. The large number of unknown fungi makes it difficult to investigate the speciation in more complex environmental studies looking into a wider diversity of species. The large speciation is organized so that the relating species are clustered together in different orders that all share the same genetic markers. With a big part of the fungal species still undiscovered the branches in the phylogenetic tree are in some part unsure and rearrangements within the tree and major branches still happen (Mora et al., 2011).

Molecular method for species identification uses differences in the nucleotide composition (molecular taxonomic units) that relate groups and species together with the similar variation in the DNA sequence (Begerow et al., 2010). Finding universal locus (barcode) for the fungal kingdom is difficult because of the broad variation in the sequence within all species. The barcode sequence can be chosen in different parts of the genome that have different functions. If the barcode region should identify sequences to species levels larger amount of differentiation is required. Pseudogenes or non-coding loci which are not transcribed have a high level of polymorphisms comparing to transcribed genes. Some of the non-coding loci have enough differentiations to identify sequences to species level. Highly polymorphic sequences can only be identified when closely related species with similar differentiations are cataloged in the database. The high level of unknown species mentioned earlier makes the database incapable to find closely related species to polymorphic

sequences due to lack of reference sequences for many species. In the highly polymorphic areas the big amount of changes has made these regions lose the similarities against further related sequences. The amount of changes is beneficial when a new sequence is compared in the database to determine the speciation. When closely related species are missing in the database the polymorphic locus is often unidentified. Parts of the genome that are highly stable and code for an important function in the cell will have low levels of differentiation. New mutations in important loci are not favored. These lower amounts of differentiations will separate species in higher taxonomy but in some cases lack the ability to match the sequence down to a species level even if these sequences are known (Dubouzet and Shinoda 1999). The ideal barcode region is constant within the same species and is unique for only that species. The loci that are used for identifying species would have to be the same for all species (Schoch et al., 2012).

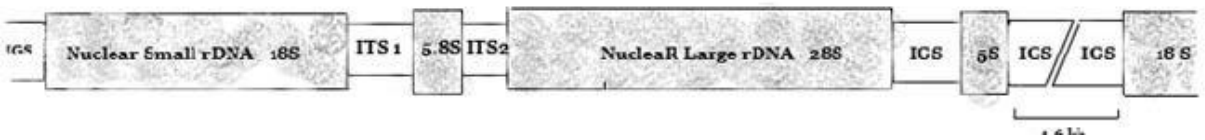


Figure 1. ITS map that contains the small rDNA 18s subunit (SSU), ITS1 region, transcribed region 5,8S, ITS2 region, the large nuclear rDNA subunit 28S (LSU).

The most common used barcode locus to classify fungi species is the internal transcribed spacer (ITS) region. The ITS in a highly polymorphic non-coding region with enough molecular taxonomic units is able to separate sequences down to a species level. It is a fast evolving region but has lost all connection to higher taxonomy, as family. The ITS sequence is placed between two coding parts of the ribosomal subunits, the 18S small subunit (SSU) and the 28S large subunit (LSU), see figure 1. ITS region is divided into two parts, ITS1 and ITS2, which are separated with a coding region of 5.8S in between. Both the LSU and the SSU regions are highly conserve sequences with few differentiations. ITS, LSU and SSU are part of the nucleolus organizer region, NORs (Heitman et al., 2007). The rDNA copy number in the cell can vary between strains from around 40 to 200 copies. In an Aspergillus species the average strain was found to have around 50 rDNA copies (Herrera et al., 2009).

Species without good matching references sequences in database when comparing the ITS region are investigated in this study. When the highly polymorphic ITS regions give poorly matches in the database they are likely to belong to an undescribed or previously not sequenced group of sequence. The high variation in the ITS region makes it impossible to use for further phylogenetic categorizing such as genus, family and order. When no closely related species are found in the database with a polymorphic sequence other more conserved part of the genome must be used to identify the sequence. The conserved regions SSU and LSU that are flanking sequences of ITS were used in this study for phylogenetic characterization of the unknown species. Together with the SSU/LSU regions the ITS

sequences made it possible to determine the high phylogenetic taxa of the undetermined sequence.

The starting samples used in this study have been collected from environmental studies previously analyzed with 454 pyrosequencing. The ITS regions are known but the species samples are complex and contain high variation of species from that study. Designed primer for the conserved individual ITS regions makes the primers highly sequence specific. With primers in the ITS region together with universal primers in the SSU and LSU region a conserved sequence for higher phylogenetic characterization can be “fished out”.

A wide range of universal primers for rDNA LSU- and SSU region are available for use. These can match together with a specific primers and give products. All primers from LR5 to LR7 are found to be working on a broad range of fungal species and the set of NS1 and NS2 are common primers that are shown to work in fungi, protists, red- and green algae (White et al., 1990). The primers used in this study can be seen in table 1.

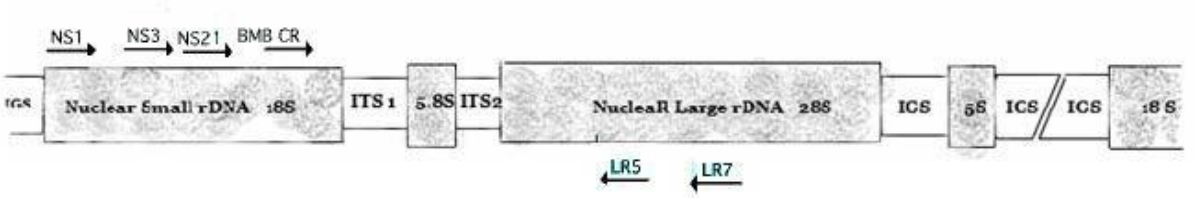


Figure 2. Primer used in the PCR amplification for both SSU and SSU region.

Material and method

PCR amplification

The specific primers were designed using Primer3 software (http://frodo.wi.mit.edu/). The PCR recipe that was used were H_2O 45%, GreenTaq buffer 20%, dNTP mix (2mM) 20%, DreamTaq polymerase (10µM) 1%, primer1 (10µM) 4%, primer2 (10µM) 4% (primer 1 and 2 see figure 2 and table 1), $MgCl_2$ (25mM) 6%.

The PCR was used with different conditions, first condition was 2 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 50°C, 90sec at 72°C and finishing with 5 min at 72°C.
The second condition was 2 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 55°C, 90 sec at 72°C and finishing with 5 min at 72°C.

Table 1. Primer used in the amplification of SSU and LSU regions, the specific primers are coding in the ITS region for the specific sequences. Primers are designed to anneal to both forward and reverse strand.

SSU, 18S small rDNA subunit		Product Size (bp)
NS 1 (forw)	GTAGTCATATGCTTGTCTC	
NS 21 (forw)	GAATAATAGAATAGGACG	802-819
NS 23 (forw)	GACTCAACACGGGGAAACTC	1184-1203
BMB CR (forw)	GTACACACCGCCCGTCG	1637-1643
LSU, 28S large rDNA subunit		
LR 5 (rew)	ATCCTGAGGGAAACTTC	997-981
LR 7 (rew)	TACTACCACCAAGATCT	1483-1467
Specific primer (ITS)		
scata1114_1099 (for)	AGTATTGCTGTTCAGCTC	
scata1114_1099 (rew)	ACAAGGCAACAACACTCT	
scata1114_346 (for)	TCTTCTAATTTGGTCTCAA	
scata1114_346 (rew)	CAGAGAACAGATGGTTTT	
scata1114_725 (for)	GCTTTAAATGCAGTAGTGT	
scata1114_725 (rew)	TCTGCCTCTGAGGACTA	
scata1079_229_Islands_Asco3 (for)	CCCCAGGAGTAGCAAAA	
scata1079_229_Islands_Asco3 (rew)	ATCCCGAGGAGAGGTTT	
scata1079_34_Islands_Asco2 (for)	ATTTTATTTCTGGGTTGA	
scata1079_34_Islands_Asco2 (rew)	GTCAACCCAGAAATAAAA	
scata1079_3_Islands_Asco1 (for)	ACCCCTCAAAGATGACTC	
scata1079_3_Islands_Asco1 (rew)	AACATTTAGAAATTCATCA	
Ho (for)	AGATTATTCCGTCTACGAG	
Ho (rew)	GCTCGTAGACGGAATAAT	

Cleaning of PCR product

To every µl PCR product 0.05 µl glycogen was added. One tenth of the new sample volume of sodium acetate was mixed with 2.5 v of 96% ethanol and mixed with the samples and incubated for 60 min in -70 °C. Then cold centrifuged at -9 °C for 10 min at 13000 rpm after the flow throw was discarded. 100 µl of 70% ethanol was added and the samples were then cold centrifuged at -9 °C for 2 min at 13000 rpm after the flow throw was discarded. 3µl was added to all samples without sample 5 witch added 10 µl.

Cloning
In the ligation reaction 2 µl of samples, 0.5 µl salt solutions and 0.5 µl TOPO vector were mixed together and incubated at roomtemperature for 30 min. After that 25 µl of competent cells per sample was used with 2 µl from the ligation reaction and incubated at ice for 30 min. The cells were heat-shocked at 42°C in water bath for 30 sec and after that at ice for 2 min. 250 µl SOC medium was added and the tubes were placed on a shaker at 150-200 rpm at 37°C for one hour. The cellmixture was distributed over X-gal and Ampicillin treated LB plates. The antibiotic resistant in the TOPO vectors and the resistant in the cells used secured that the cells with a vector could be selected. The plates were incubated at 37°C over night.

PCR amplification and sequencing
The white/blue colonies were picked and transferred to 150µl of water. The insert was the amplified it a PCR reaction. The PCR conditions were 2 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 45°C, 90 sec at 72°C and finishing with 5 min at 72°C using the M13 forward and M13 rowers primes.

The sequenced sample was investigated with blast search (GeneBank) and a phylogenetic tree was constructed from all subgroups in the order of Ascomycota fungi (see supplementary 1).

Result
No amplification of the SSU region for any of the samples was made. The LSU region sample H0 was successfully amplified and sequenced, see figure 2. The rest of the samples were not successfully amplified in the PCR reaction.

The alignment of H0 ITS region together with H0 SSU showed no differentiation between the original ITS sequence and the sequenced sample. Both the primers LR7 and the specific designed primer were found and the M13 vector could be cut off.

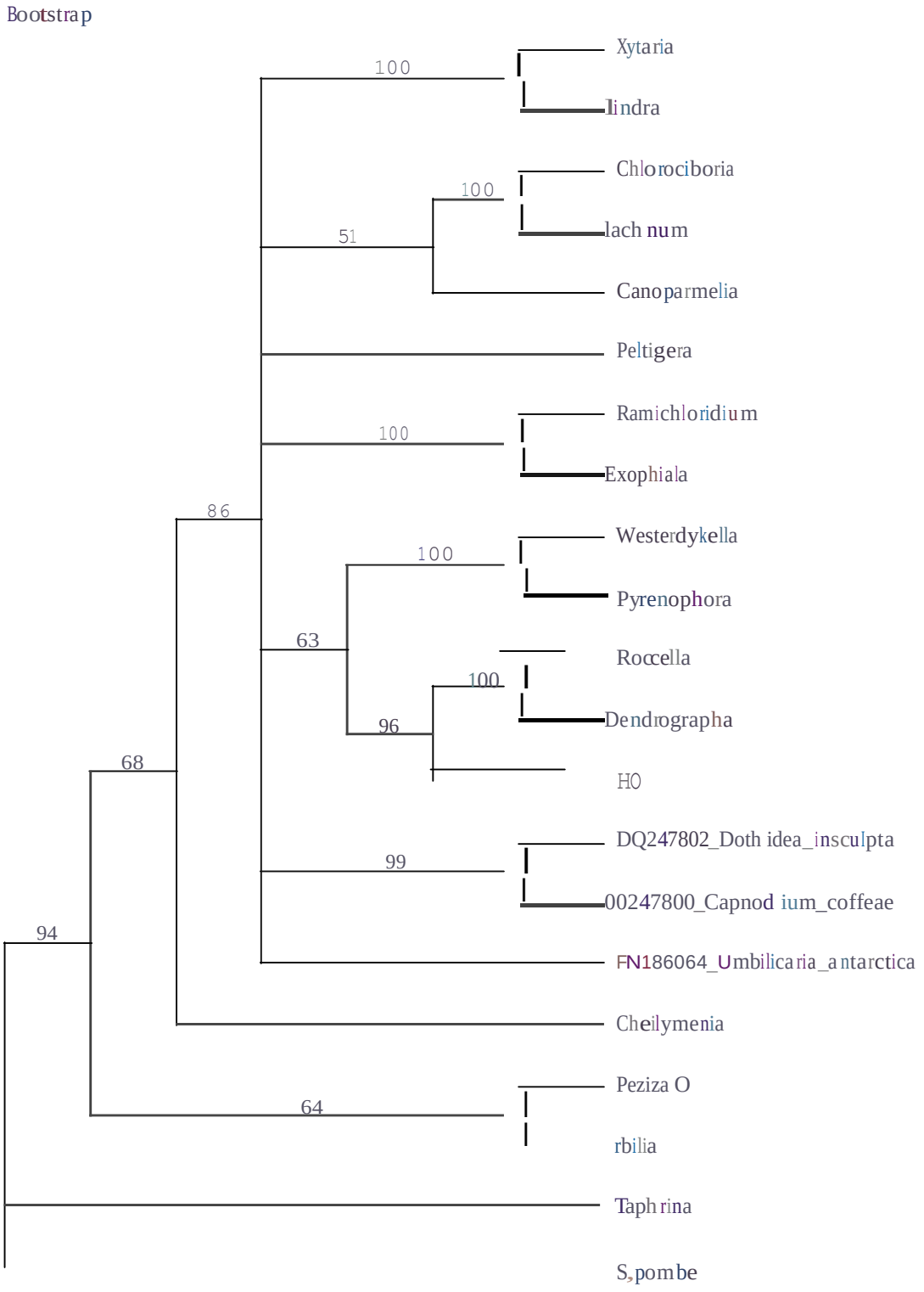
The blast search with the H0 sequence showed a low max identity on 89% with the LSU sequence. That confirms the hypotheses that the sample belongs to a new species. To be able to further investigate the species phylogenetic relations both neighbor joining and maximum parsimony analysis were done. Species from Sordariomycetes, Leotiomycetes, Lecanoromycetes, Eurotiomycetes, Dothideomycetes, Arthoniomycetes, Pezizomycetes and Orbiliomycetes were used to cover all orders of Ascomycota. As out-group two species of Taphrinomycotina were used which according to the blast search were distant to the H0 sample, see figure 3 & 4. Both neighbor joining and maximum parsimony analysis sorted the H0 sample as a side group of Arthoniales with high bootstrap value on 96 and 98

Discussion
The study aimed to sequence a conserved region of the rDNA sequence from a complex community sample instead of a non-cultivated sample. The LSU region was successfully done with on sample, H0.

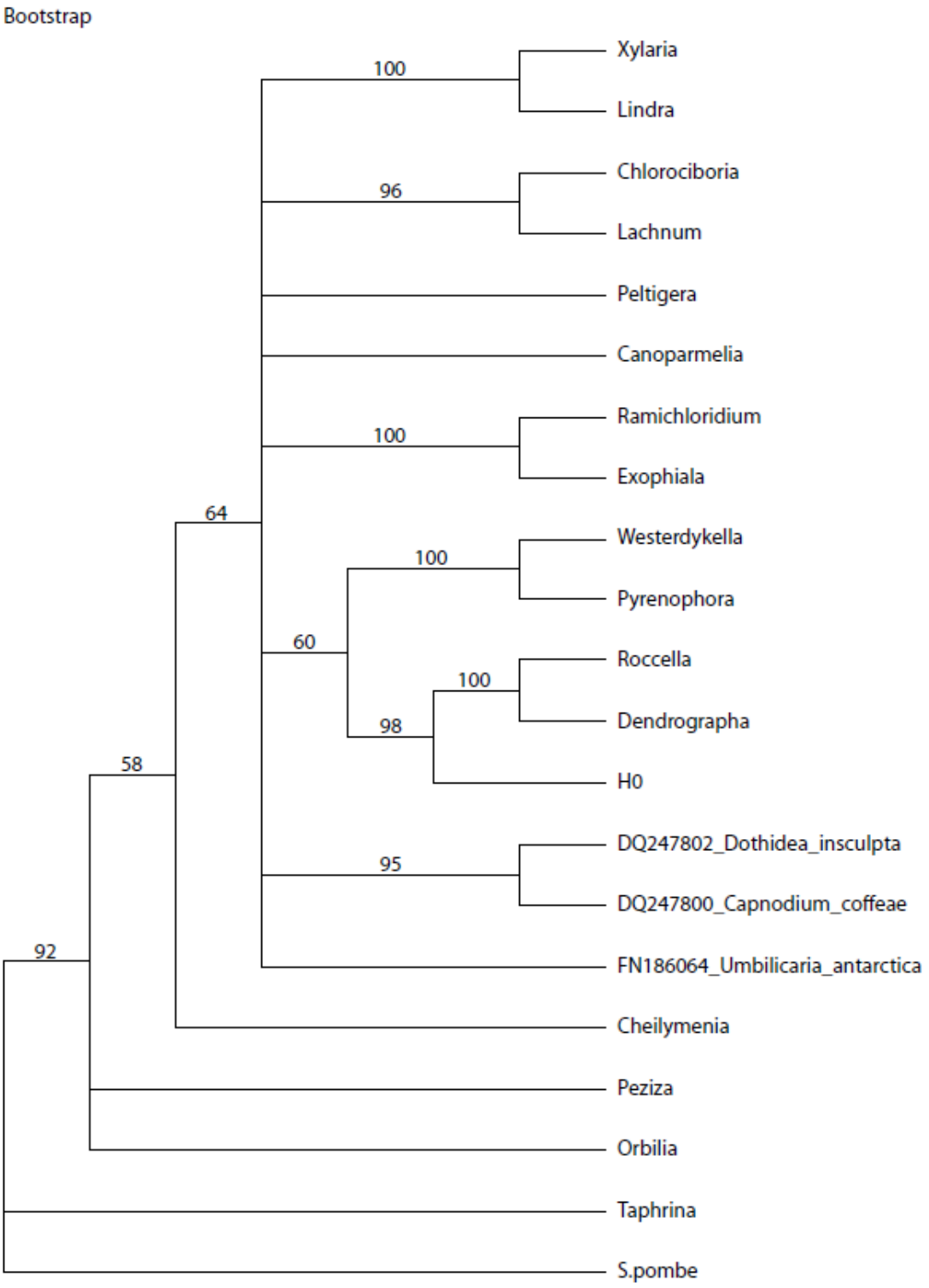
The overall problem in this study was the PCR amplification of LSU and SSU which are placed before and after the ITS region of the ribosomal DNA sequence. The lack of amplifications success was most likely caused by errors in the designed primers. The primer design was done with the software Primer3 invented by the National Institutes of Health. The primer design software was set on around 18 nucleotides long with the same annealing temperature as the LR5 and NS1 primers. Several sets of specified markers in parallel PCR amplifications would have higher possibility to amplify the right product. That would also save some time which was the limiting factor for this study. The biggest problem was to combine the universal primer with the specific designed primer to work at the same annealing temperature. Factors as, concentration of primers, concentration of $MgCl_2$, concentration of polymerase were tested without any result.

Table 2. SSU sequence from sample Ho, aligned with primer LR7, the specific primer for Ho and Ho ITS sequence.

```
Sample bp sequence

H0:     0´ TACTACCACCAAGATCTGCACTAGAGGCCGTTCTACCCGGGATCACTCCCTAGGCTTCGTCACGGGCCTCCACGCCTGCCTACAC
LR7:       TACTACCACCAAGATCT---------------------------------------------------------------------

H0:    85´ CGGGGCTTAGTATCTGCCCCGGCGGCGGGGTATAGGTAATACGCTTGAGCGCCATCCATTTTCAGGGCTATAACGTTCGGCAGGT

H0:   170´ GAGTTGTTCACACTCCTTAGCGGATTCCGACTTCCATGGCCACCGTCCTGCTGTCTAAACGTTATAACACCTTTTGTGGTGTCTG

H0:   255´ ATGAGCGTACATTCCGGCACCTTAACCCCGCTTTCGGTTCATCCCGCATCGCCAGTTCTGCTTACCAAAAATGGCCCACTAGTAA

H0:   340´ CGGTGCATTCAAATGTCCGCGTTCAATTAAGTAACAAGGACTTCTTACATATTTAAAGTTTGAGAATTGGTTAAGGTTGTTTCAA

H0:   425´ CCCCAAGGCCACTAATCATTCGCTTTACCTCATAAAACTGAAAACGTTACTGCTATCCTGAGGGAAACTTCGGCAGGAACCAGCT

H0:   510´ ACTAGATGGTTCGATTAGTCTTTCGCCCCTATACCCGAATTTGACGATCGATTTGCACGTCAGAA-CC-GCTGCGAGCCTCCACC

H0:   595´ AGAGTTTCCTCTGGCTTCACCCTATTTAGGTATAGTTCACCATCTTTCGGGTCCCAACAGCTATGCTCTTACTCAAATCCATCCG

H0:   680´ AAGACATCAGGATCGGTCGATGGTGCACCTTGCGGTTCCCACCTCCGTTCACTTTCATTACGCGCATGGGTTTTACACCCAAACA

H0:   765´ CTCGCATAGATGTTAGACTCCTTGGTCCGTGTTTCAAGACGGGTCGCTTGCAGCCATTACGCCAGCATCCTAGCAGAGCGCGGTC

H0:   850´ CTCAGTCCCGGGTGGCCGCATGATGCCAGAGGCTATAACACTCCCCCCCGCGG-AAGAATACACAGCCCAGAAGGACTGCTGAAC

H0:   935´ GAAGCGGAGCCACATTCCCCTGGCGTTTATCCGGCCCCC-GAAACTGATGCTGG-CCTGTGACCGCGGCCAAGTCTGGCTGCAAG

H0:  1020´ AGCTTCCCTTTCAACAATTTCACGTGCTTTTTAACTCTCTTTCCAAAGTGCTTTTCATCTTTCGATCACTCTACTTGTGCGCTAT

H0:  1105´ CGGTCTTTGGCCGGTATTTAGCTTTAGAAGAAATTTACCTCCCATTTAGAGCTGCATTCCCAAACAACTCGACTCTTTGAGTGGG

H0:  1190´ TTCCACACGGCAAAGGCATCCCGAACGAAGACGGGGTTCTCACCCTCTATGACGTCCTGTTCCAAGGAACCTAGCACGGGCCCAA

H0:  1275´ TGCCGGAACACCCTCTGCAAATTACAACTCGGGACCCGAAGGGACCAGATTTCAAATTTGAGCTATTGCCGCTTCACTCGCCGTT

H0:  1360´ ACTAGGGCAATCCCAGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTATCCCTACCTGATCCGAGGTCAA
ITS:       ------------------------------------------------TTAAGTTCAGCGGGTATCCCTACCTGATCCGAGGTCAA

H0:  1445´ TGTCTAGGTAAAAATTTGGTTGTCTGGAAGCGGGCGGTCGCCGGCGCTCGTAGACGGAATAATCT-------------------
ITS:       TGTCTAGGTAAAAATTTGGTTGTCTGGAAGCGGGCGGTCGCCGGCGCTCGTAGACGGAATAATCTACTACGCTCGAGGCCTTCGA
Specific:  ------------------------------------------------CTCGTAGACGGAATAATCT-------------------

ITS:       CCACCGCCGACGCATTTAGGGCCCGCCCCCTGTGACGGGGGCGAGGGGCCCAAGGCCAAGCAAAGTGCTTGATTTTT-------
```



Figur 3. Neighbor joining analyses for HO,were Xylaria acuta and lindra thalassiae belongs to Sordariomycetes, Chlorociboria and Lachnum belongs to Leotiomycetes,Peltigera and Canoparmelia belongs to Lecanoromycetes, Ramichloridium and Exophiala belongs to Eurotiomycetes,Westerdykella and Pyrenophora and Dothidea insculpta and Capnodium coffeae belongs to Dothideomycetes,Roccella and Dendrographa belongs to Arthoniomycetes, Cheilymenia and Peziza belongs to Pezizomycetes,Orbilia belongs to Orbiliomycetes,two out-groups of Taphrina and S.pombe belonging to Taphrinomycotina. HOare groping along with Arthoniales with a bootstrap value on 96.



Figur 4. Maximum parsimony analyses for H0, were Xylaria acuta and Lindra thalassiae belongs to Sordariomycetes, Chlorociboria and Lachnum belongs to Leotiomycetes, Peltigera and Canoparmelia belongs to Lecanoromycetes, Ramichloridium and Exophiala belongs to Eurotiomycetes, Westerdykella and Pyrenophora and Dothidea insculpta and Capnodium coffeae belongs to Dothideomycetes, Roccella and Dendrographa belongs to Arthoniomycetes, Cheilymenia and Peziza belongs to Pezizomycetes, Orbilia belongs to Orbiliomycetes, two out-groups of Taphrina and S.pombe belonging to Taphrinomycotina. H0 are groping along with Arthoniales with a bootstrap value on 98.

The samples in this study were taken from environmental studies and were diluted to lower the concentration of inhibitors. The set of samples scata1114_1099, scata1114_346, scata1114_725 were all take from the same study and did not show any result from the sets of different PCR attempts. The samples were gathered from other researches and were already diluted 100 times. It was never tested if another concentration would have made the degree of PCR amplification higher or if the annealing temperature for the different primers caused this problem. Dilutions for up to 500 times might have lower the inhibitors in a more efficient way that the templates could have been able to been amplified, but this was not tested.

The H0 sample with the LSU region which contained a partial ITS region showed 100% similarity with the initial ITS sequence for the overlapping area. That proves that the LSU sequence and ITS are from same species. In the blast a smaller sequence after the ITS was used which gave high quality matches. The algorithm in the blast search compares the whole sequence and gives a rate of similarity. That favors longer sequences to get a better score. The LSU sequence in GeneBank mostly contains partial LSU and SSU sequences and a consideration must be done if the analyzed sequence should contain high variation or a region that has lots of corresponding hits in the database (personal communication Björn Lindahl).

That is an indication for a new branch in the order of Arthoniales that has not been documented before. Arthoniales is a small order which only contains lichens species. The H0 sample in this study was collected from needles. It is unlikely that lichens species would have grown on needles when that is not their natural environment. Further analyzing showed that the support for the Arthoniales branch in the tree of Ascomycota species is divided and has a side branch which belongs to Dothideomycetes species. Both the neighbor joining and maximum parsimony trees had members from this Dothideomycetes branch without affecting the bootstrap value of Arthoniales and H0. The top hits in the blast search all showed different species in the order of Dothideomycetes. The H0 sample is phylogenetic like the Arthoniales species but most likely it dose not belong to that order. The support for the Arthoniales is weak when the branch has species both in Arthoniales and Dothideomycetes. Sister group of Arthoniales could belong to something totally different and not for certain lichens species.

Conclusions from this study are that the LSU locus in the ribosomal DNA has a lower evolving rate. That makes this locus able to categorize sequence in a higher order (up to class in this study) than the ITS region. The ITS regions did not have a marching hit in the Genebank databank, which indicates that it could belong to an undescribed species. The National Istiuonal of Health launched the database GeneBank is the common tool for species identification. Today over 2.4 million fungal sequences are available in the database to be used in the blast search (Begerow et al., 2010). The database limitations are the biggest problem for using ITS as a barcode system. The ITS region is fast evolving and over time the probability for sequential substitutions is getting higher. With a high sequential substation rate then the new mutations still have a higher chance to keep the same degree of

divergence. The actual number of differences between two sequences is then lower than in reality which makes it hard to find evolutionary similarities (Graur and Li, 2000) in the GeneBank database. But the high degree of divergence in the ITS region is the factor that makes it outstanding for species identification comparing with SSU and LSU regions. The interspecific variation within the ITS locus for all species makes it a useful standard barcode system (Schoch et al., 2012).

Studies on the SSU region with Schoch et al, showed that SSU have low levels of interspecific variation compared to the variation with other sequences. This makes the SSU region too conserved for being a useful barcode locus (Schoch et al., 2012). Results from studies from Bruns et al, which investigated the OTU in the SSU locus, investigated the relationship within the fungi kingdom. Four divisions were investigated, Chytridiomycota, Zygomycota, Ascomycota and Asidiomycota. Between the four OTU 99% of the sequences had fewer than eight changes. A low amount of species had four or more changes while the minority had less than or zero changes. The differences in OTU in the Ascomycota sequences showed that 99% of the sequences had lower than four changes. In this alignment an even higher rate of sequences had no changes and small part of the sequences with over three changes (Bruns et al., 1992). The highly conserved SSU region has few hypervariable units to separate related species and demands closely related sequences for comparison in the database. That limits the use for SSU as a universal barcode locus.

Conclusions

The broad genetic diversity in the fungal kingdom has been formed over a long evolutionary time (Schoch et al., 2012). The attempt to cover all fungal species may not be the goal with the Genebank database. Some fungal species are fast evolving and new species lines forms continually. To be able to fast categorize sequences with as much data as possible universal barcode locus is a good tool. Using a polymorphic region which has the possibility to fast give lots of and trustful information has heavily reduced amount of relative species that could be categorized in the database. A conserved region gives less information but the amounts of relative species are broad which could be matched in the Genebank database. The complex problem to find a universal barcode will be difficult to work in the same rate over all fungi species. The consensus is to find one or a few locus that can work for a broader amount of species. With the knowledge of their limitations they can be used in a useful way. ITS have been found to be easy to amplify and is easy to fast analyze which works well with a lot of species. LSU region combined with ITS showed over 100 operational taxonomic units (Rosling et al., 2011). That would make LSU combined with ITS a useful universal barcode system for fungi species.

This study showed some evidence for a new sister group of Arthoniales. The position of this new group could not be determined if it belongs to Arthoniales or Dothideomycetes. Most likely it is a group of Dothideomycetes but that could not be proven.

Acknowledge

This research was supervised by Lindahl, Björn .Clemmensen, Karina Engelbrecht provided samples and ITS sequences. Millberg, Hanna and Ihrmark, Katarina furnished technical assistance. Proofreading by Simin Cai.

References

BEGEROW, D., NILSSON, H., UNTERSEHER, M. & MAIER, W. 2010. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Applied Microbiology and Biotechnology,* 87**,** 99-108.

BRUNS, T. D., VILGALYS, R., BARNS, S. M., GONZALEZ, D., HIBBETT, D. S., LANE, D. J., SIMON, L., STICKEL, S., SZARO, T. M., WEISBURG, W. G. & SOGIN, M. L. 1992. Evolutionary Relationships within the Fungi: Analyses of Nuclear Small Subunit rRNA Sequences. *Molecular Phylogenetics and Evolution,* 1**,** 231-241.

GRAUR, D. & LI, W. H. 2000. *Fundamentals of Molecular Evolution*, Sinauer.

HEITMAN, J., KRONSTAD, J. W., TAYLOR, J. W. & CASSELTON, L. A. 2007. Sex in Fungi Molecular Determination and Evolutionary Implications Preface. *Sex in Fungi: Molecular Determination and Evolutionary Implications***,** Xv-Xvii.

HERRERA, M. L., VALLOR, A. C., GELFOND, J. A., PATTERSON, T. F. & WICKES, B. L. 2009. Strain-Dependent Variation in 18S Ribosomal DNA Copy Numbers in Aspergillus fumigatus. *Journal of Clinical Microbiology,* 47**,** 1325-1332.

MORA, C., TITTENSOR, D. P., ADL, S., SIMPSON, A. G. B. & WORM, B. 2011. How Many Species Are There on Earth and in the Ocean? *Plos Biology,* 9.

ROSLING, A., COX, F., CRUZ-MARTINEZ, K., IHRMARK, K., GRELET, G. A., LINDAHL, B. D., MENKIS, A. & JAMES, T. Y. 2011. Archaeorhizomycetes: Unearthing an Ancient Class of Ubiquitous Soil Fungi. *Science,* 333**,** 876-879.

SCHOCH, C. L., SEIFERT, K. A., HUHNDORF, S., ROBERT, V., SPOUGE, J. L., LEVESQUE, C. A., CHEN, W. & CONSORTIUM, F. B. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 6241-6246.

WHITE, T. J., BRUNS, T., LEE, S. & TAYLOR, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenies. *PCR Protocols: a Guide to Methods and Applications (eds Innis MA,Gelfand DH, Sninsky JJ, White TJ).* San Diego, California: Academic Press.

Supplementary 1. ITS sequences from samples scata1114_1099, scata1114_346, scata1114_725, scata1079_229_Islands_Asco3, scata1079_34_Islands_Asco2, scata1079_3_Islands_Asco1, Ho + NS1

>scata1114_1099 ITS sequence
TTAAGTTCAGCGGGTTTTCTTATGCTGCAGAGGTCAGCAAGAAGAGAGTGTTGGAGCGACGC
AAACAGCATCCCGTGTTGATTCCTGCCGCACCGAAAAGGCAAGGCAGGCTCGACGGGCGAGC
TTGCGCAGATCTCAGAGCTGAACAGCAATACTGCTGCTCAACACCGAATCCTTTCGGGACAA
AGCCGTAAGATGAATTGCACAAGGCAACAACACTCTGCAGCACATGGCCGGCCTGGGAAAAT
GCCGGACCGCAATTTGCATT

>scata1114_346 ITS sequence
TTAAGTTCAGCGGGTAGTCCTGCCTGATTTGAGACCAAATTAGAAGATATAGTATTGTGTCC
ACAGGCAGAGACGCCTGCTGCGACCAAACACAAGAATGTGTTTATAGCACACAGCAGGGCCT
CCATCAGCGGCACCTTCTTGCTGCTTTCCAGGCCTTGCATACACGATAGCACATGCAGAGGC
CCAATTCGCCCTCTTTGCAGAGAACAGATGGTTTTTGCGGTACTCAAACAGACATGCCCATT
GCTGGGCGCAAGTTGCGTT

>scata1114_725 ITS sequence
TTAAGTTCGGCGGGTAGTCTTACGTGATTTGAGATCAGTTGCTGTGTACACGCTTTTATTAT
TATGCGCACACACGTTTACGAGTGCACTTGCTTCTGCGCATACGCAAGGGGACGCACCGTGA
GTAGACACCAAGATCTCCAGCACACAGTGTCCACCAAGCGCGAGTCGCGCAGAGATACACTA
CTGCATTTAAAGCGCTCTGGGCGCTCAAGTCTGCCTCTGAGGACTAGAGACAAGATACTGTT
ACGATTCTCAAACACGTATACCCTCCATGAGAGGGTGCAAGGTGCGTT

>scata1079_229_Islands_Asco3 ITS sequence
TTAAGTTCAGCGGGTATCCCTACCTGATCCGAGGTCAAGACTTGAAGAATGGGTGTGTTCCCAGGCGG
GCGAGGCACCGGATCCCGAGGAGAGGTTTTTTGCTACTCCTGGGGCCGGTGCCGCGCCGCCTACGAGT
TTCGGGCACGTCCGCAGCGGGGACGAGGCCCAACACCAAGCGAGGCTTGAGGGTTGAGATGACGCTCG
AACAGGCATGCCCACCGGAGTACCGGAGGGCGCAATGTGCGTT

>scata1079_34_Islands_Asco2 ITS sequence
TTAAGTTCAGCGGGTATCCCTACCTGATCCGAGGTCAACCCAGAAATAAAATTGGGGTTGAGAGGCAG
ACATTTATCTGACAGAAGCGATGAAGTTTCCTAGACGCTTGAAGCCAGGCAGTGTCGCCATTGCATTT
GAGGCGTGTCCTCCAGAGAGGACAACACCCAAGACCAGGCCTACAAGTAGGCTTGAGTTTAGCAAATG
ACGCTCGAACAGGCATGCCCTCCGGAATACCAGAGGGCGCAATGTGCGTT

>scata1079_3_Islands_Asco1 ITS sequence
TTAAGTTCAGCGGGTATTCCTACCTGATCCGAGGTCAACATTTAGAAATTCATCAATTGACTGTCAAC
ATCCCGCCAGCTAGACGACGTGTGTGCGCCTGTGGTTGAATATCACCATCGCTGATTTTGAGGCATGC
CTTTAGGCATTGCCCAACGCCAAAGAGTCATCTTTGAGGGGTGTAATGACGCTCGAACAGGTATGCCC
TTCGGAATGCCAAAGGGCGCAATATGCGTT

>Ho ITS sequence
TTAAGTTCAGCGGGTATCCCTACCTGATCCGAGGTCAATGTCTAGGTAAAAATTTGGTTGTCTGGAAG
CGGGCGGTCGCCGGCGCTCGTAGACGGAATAATCTACTACGCTCGAGGCCTTCGACCACCGCCGACGC
ATTTAGGGCCCGCCCCCTGTGACGGGGGCGAGGGGCCCAAGGCCAAGCAAAGTGCTTGATTTTT