5-19-2010

Phylogenetic Analysis of Aspergillus and Related Fungi Based Mitochondrial Cytochrome Oxidase 1

Shaunte Hulett-Abdin

PHYLOGENETIC ANALYSIS OF ASPERGILLUS AND RELATED FUNGI BASED ON MITOCHONDRIAL CYTOCHROME OXIDASE I

THESIS

SHAUNTÉ HULETT-ABDIN

2010

PHYLOGENETIC ANALYSIS OF ASPERGILLUS AND RELATED FUNGI BASED ON MITOCHONDRIAL CYTOCHROME OXIDASE I

THESIS

Presented in Partial Fulfillment of the Requirements for

the Master of Science Degree in the Graduate School

of Texas Southern University

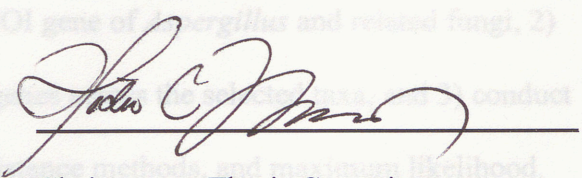
By

Shaunté Hulett-Abdin, B.S.

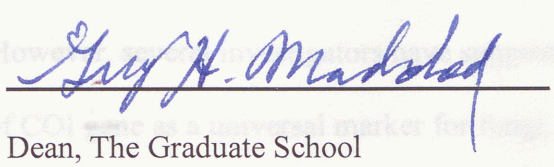
Texas Southern University

2010

Approved By



Chairperson, Thesis Committee



Dean, The Graduate School

PHLYOGENETIC ANALYSIS OF ASPERGILLUS AND RELATED FUNGI BASED ON MITOCHONDRIAL CYTOCHROME OXIDASE I

By

Shaunte' Hulett-Abdin, M.S.

Texas Southern University, 2010

Professor Hector Miranda, Advisor

Microorganisms are hypothesized to experience rapid rates of mutation in space due to microgravity and ionizing radiation. As an initial effort to test the relative rates to which certain genes evolve under space-like conditions, we reviewed the usefulness and reliability of mitochondrial cytochrome oxidase 1 (COI) gene in resolving the evolutionary relationships of *Aspergillus* and related fungi. The objectives of this study are to 1) characterize the mitochondrial COI gene of *Aspergillus* and related fungi, 2) align the homologous sequences of COI genes across the selected taxa, and 3) conduct phylogenetic analysis using parsimony, distance methods, and maximum likelihood.

Most recently, the 600 bp region of the 5'end of the COI gene was proposed to be the ideal marker for barcoding eukaryotes. However, several investigators have suggested potential problems associated with the use of COI gene as a universal marker for fungi. Of those, they include failure to resolve closely related taxa and extensive overlap of

intra- and interspecific diversity. The universality of the gene is also questioned by reports of the prevalence of mobile introns across many fungal groups. We investigated the utility of COI gene in discrimination between species, between strains and between geographic isolates.

We sequenced 5 species from ATCC, plus an unknown culture obtained from environmental samples. For the analyses, alignments were done using Geneious Pro (Biomatters, Ltd., New Zealand). Final alignments were exported in nexus file format. Using PAUP©4b10, we conducted phylogenetic analyses thru Maximum Parsimony (MP), distance method using Neighbor-Joining, and Maximum Likelihood (ML).

Support for nodes were estimated using nonparametric bootstrap and were done for MP, distance and ML analyses. The substitution model of molecular evolution used was Jukes-Cantor.

Our preliminary analysis suggested the following observations. 1) The conflict in the utility of COI as barcoding marker reflects the lack of taxonomic stability within Fungi at the intra- and inter- species level. There is a need to expand geographic sampling of taxon to further test the validity of current classification. 2) No intron was observed within the short barcode region (600 bp) among the samples we investigated.

Approved By

	5/3/2010
Chairperson, Thesis Committee	Date
	5/3/2010
Committee Member	Date
	5/3/2010
Committee Member	Date
	5/3/2010
Committee Member	Date

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

ACG	Area Conservation de Guanacaste
ATCC	American Type Culture Collection
BenA	Beta Tubulin
BLAST	Basic Local Alignment Search Tool
β-tubulin	Beta Tubulin
bp	Base Pair
COI	Cytochrome Oxidase 1
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
ITS	Internal Transcribed Spacer
MD1	Microbial DNA Solution 1
MD2	Microbial DNA Solution 2
MD3	Microbial DNA Solution 3
MD4	Microbial DNA Solution 4
MD5	Microbial DNA Solution 5
ML	Maximum Likelihood
MP	Maximum Parsimony
mtDNA	Mitochondrial DNA
NADH	Nicotinamide Adenine Dinucleotide
ND6	Dehydrogenase Subunit 6
PAUP	Phylogenetic Analysis Using Parsimony
PCR	Polymerase Chain Reaction

RNA	Ribonucleic Acid
SDS	Sodium Dodecyl Sulfate
TAE	Tris-Acetate-EDTA

VITA

September 11, 1978	Born – Flowood, Mississippi
1998-1999 & 2001-2002.....	Lead Teacher Agapé Christian Academy Clinton, Mississippi
1999-2001	Assistant Teacher Clinton Public Schools Clinton, Mississippi
2002-2004	Assistant Appeals Coordinator Health Information Designs Flowood, Mississippi
2007	B.S., Texas Southern University Houston, Texas
2008-2010	Laboratory Coordinator Department of Biology Texas Southern University
Major Field	Biology

ACKNOWLEDEMENTS

I would like to first thank my mentor, **Dr. Hector Miranda**, for his dedication to my graduate education. Dr. Miranda was never more than a phone call or e-mail away. He accepted me into his lab without haste and immediately put me to work. I enjoyed working in his phylogenetics lab and appreciate the skill set I am leaving with. Thank you for guiding me through my graduate studies.

Next, I have to acknowledge my Principle Investigator and committee member, **Dr. Olufisayo Jejelowo**. Dr. Jejelowo has provided me with many leadership, research, and career opportunities within the department of biology and NASA-CBER URC. I am truly thankful for her sincere interest in my advancement. I am quite fortunate and full of gratitude, Dr. Jejelowo, for your encouragement, vision, strength, diligence, and love.

To my other committee members, **Dr. Jason Rosenzweig** and **Dr. Omonike Olaleye**, thank you for taking on the task of guiding me through my thesis work. Your direction is greatly appreciated.

To **Dr. Desireé Jackson** and **Dr. James DuMond**, my unofficial advisors, your doors have always been open to me. Over the years you all have seen me at my best and at my worst. I appreciate the tough love you both showed while never leaving my corner. I will forever be grateful to you both.

To my mom away from home, **Mrs. Helen Cockrell**, there is absolutely no way I can say thank you enough for all that you have done and are to me. You are my blessing and I thank you for always giving the right advice. Your mom instincts kick in, and you know when I need a word from "momma". Thank you.

To my lab mate, **Ms. Tiarra Spencer**, thank you for assisting me with whatever I needed in the lab. You are a talented mind, and I feel honored to have worked with you.

To **Ms. Cherita Thomas**, thank you for adopting me as your graduate "little sister". When I needed to know how to do things the correct way, I called you. You have a calm spirit and were always willing to help where you could.

To my mother, **Dorothy Hulett**, and sisters, **Christina & Jansa Hulett**, I love you all and appreciate all of the support throughout the years. We have come a long way.

To my sons **Markus & Shaun**, mommy loves you guys so much. Thank you for making being a mom so easy.

Most importantly, thank you to my husband **Ali Abdin**. You have sacrificed so much so that I could achieve my goals. Without your support, I would not be at this point. I love you.

CHAPTER 1

INTRODUCTION

Phylogenetic analysis is essential to accurately identify of the fungal species. These results yield a wealth of information, including the relatedness of varying species. The taxa, *Aspergillus* and other related fungi, have been traditionally identified based on stringent taxonomic guidelines based solely on morphology. More recently, biodiversity studies on DNA sequences provided a significant advance to fungal taxonomy and systematics (Seifert *et al.*, 2007). As a consequence, molecular phylogenetics has uncovered many cryptic species in a number of taxa. Modern molecular technologies have come with the highly developed bioinformatics infrastructure that allows ready access and exchange of data on a worldwide scale.

There exists a need for economical, quick, and precise methods of identifying fungi on a global scale that will be widely acceptable. Although, there are about 1.5 million species of fungi estimated to exist, less than 10% have been formally described (Hawksworth, 1991, 2001, 2004). The overwhelming majority of fungi are probably undescribed. The case for a DNA-based identification system is evident. DNA barcoding has been suggested as a possible identification system for fungal species (Seifert *et al.*, 2007; Min *et al.*, 2007: Kubicek *et al.*, 2008). The most popular molecular marker in fungal identification and phylogenetic reconstruction, at present, is the Intergenic Transcribed Spacer (ITS) region (Seifert 2009). However, it is difficult to

align, which restricts its use for phylogeny reconstruction. The most compelling reason to look for additional markers is that a phylogeny based on a single gene or region may actually track the evolution of the gene and not the species or lineage (Hillis *et al.* 1996). Thus, a search for markers within the fungal genome that possess resolving power to discriminate within and among species and augment the utility of the ITS region is important.

DNA Barcoding

DNA barcoding is based on a relatively simple concept that effective, broad-spectrum identification systems can be based on sequence diversity in short regions of DNA (Hebert *et al.*, 2003; Blaxter *et al.*, 2005; Savolainen *et al*, 2005). Once a barcode can be recognized as an accurate identifier for a species, it can be compiled in a public bioinformatic library. This would provide a "one-stop-shop" for identifying species in a cost effective manner. Certain criteria should be met when deciding on a DNA sequence as an identifier. DNA sequences should be orthologous in the examined organisms, and variable enough to allow species identification, with low levels of intraspecific variation (Hebert *et al.*, 2003). A DNA barcode should be easily accessible, relatively short, simple to sequence, and easily alignable with no recombination (Geiser, *et al.*, 2007). As an identification system for animal life, the mitochondrial gene cytochrome c oxidase 1 (COI) was shown to be highly effective (Hebert *et al.*, 2003), and it delivered species-level resolution in more than 95% of taxa from test sets of different animal lineages (Ward *et al.*, 2005 & Smith *et al.*, 2005). In response to those initially promising results, large interest in assessing the utility of this 648 bp region with other organisms was generated. Scientists have tested its usefulness in resolving divergence amongst species

of fish (Ward *et al.*, 2005), various arthropods (Smith *et al.*, 2005), birds (Hebert *et al.*, 2003), mammals (Hajibabaei *et al.*, 2007), algae (Saunders *et al.,* 2005), and most recently fungi (Seifert *et al.*, 2007). COI, however, failed to resolve taxonomic issues within most plant species due to the slow evolving rate in higher plants (Kress *et al.*, 2005). While initial studies have been published regarding the effectiveness of COI as a species identifier with the fungus *Penicillium* (Geiser *et al.*, 2007), its cousin, *Aspergillus,* has not yet been adequately evaluated using this same gene.

Aspergillus

Little information is available regarding complete fungal mitochondrial genomes. Scientists are diligently sequencing mitochondrial DNA (mtDNA) of fungus in the hope of making the sequences available in genetic databases. Consequently, recent studies have described the difficulty in doing so due to the presence of mobile introns (Santamaria *et al.*, 2009).

We examined various species from the genus *Aspergillus* (Trichocomaceae, Eurotiales, Ascomycetes, Ascomycota). This genus is part of a phylum of monophyletic filamentous fungi more commonly known as mold and reproduces asexually. *Aspergilli* possess immense ecological and metabolic diversity. The fungus aids in pharmacological studies as well as agricultural and food industries (Bennett, 2010). Clinical reports have also found *Aspergillus* to be pathogenic. Its virulence varies among species, and the diseases that it causes, defined as aspergillosis, range from mild allergic symptoms to life-threatening illnesses (http://www.aspergillus.org). Currently around 250 molds found across the globe in various environments are attributed to the diagnosis of aspergillosis (Bennett, 2010 & http://www.aspergillus.org).

Multilocus and single locus phylogenetic studies have been conducted in *Aspergillus* (Geiser *et al.*, 2007). The mtDNA genome was not included in those investigative studies. Regions of the nuclear genome investigated include, ribosomal RNA and β-tubulin, and they were found to suffer from recombination of genetic information (Geiser *et al.*, 2007). Mitochondrial genes are not susceptible to these problems. Given the diversity of the phenotypes of this genus, which may actually outnumber that which is uncovered by sequence analysis, a standard mode of identification is necessary. As a result, *Aspergillus* is a prime candidate for investigating the efficacy of COI as a barcode.

I conducted a pilot analysis of COI sequences from various *Aspergillus* and related fungal species BLASTed from GenBank, *Aspergillus* species cultured in the lab and an unknown isolate collected from the ventilation system at Texas Southern University. The amplified sequences were aligned and pairwise sequence divergences were estimated. Finally, phylogenetic trees were constructed using parsimony, distance methods, and maximum likelihood.

CHAPTER 2

RELATED LITERATURE

Investigations

One case study examined the effectiveness of barcode markers in Fungi. Santamaria *et al.,* expressed a need to develop a cost-effective identification system for fungal species (Santamaria *et al.*, 2009). They looked at the incidence of mobile introns said to have a negative effect on resolving divergence issues within the mitochondrial genome. The group was able to ascertain that a bioinformatic analysis of Ascomycota was very accurate (Santamaria, *et al.*, 2009). Although the mitochondrial genome is peppered with introns, differentiation of species was still possible. The full Nicotinamide Adenine Dinucleotide (NADH) dehydrogenase subunit 6 (ND6) displayed distinguishable polycistronic regions. Thus literature suggests the mitochondrial gene ND6 as a potential barcoding candidate.

In another study, *Penicillium* was the test case for the investigation, where the utility of COI as the primary barcode region for *Penicillium* was tested (Seifert et al., 2007). The sequence patterns of many fungal taxa were examined for differences in the alignment. These initial results showed COI to be effective in resolving divergence questions (Seifert *et al.*, 2007). Primers were then designed to amplify COI sequences for isolates from the family Trichocomaceae. Although the group anticipated issues with mobile introns, they determined that is was not a problem in the assessment of the experiment. Only 2 of the 370 *Penicillium* strains tested contained introns. Out of 58

species, 38 formed cohesive assemblages with distinct COI sequences sharing involving known species complexes (Seifert *et al.*, 2007). The intraspecific divergence of the *Penicillium* species averaged 0.06%, less than the two other loci tested, Internal Transcribed Spacer (ITS) and Beta Tubulin (BenA). In addition to those findings, the interspecific divergence was approximately 5.6% which could be compared to ITS but less than BenA at 14.4%. While BenA provides greater taxonomic resolution between species, CO1 requires a simpler amplification and alignment process.

Geiser *et al.* scrutinized various taxonomic practices for identification of *Aspergillus* (Geiser *et al.*, 2007). Physiological characters often show differences that reflect phylogenetic species boundaries, and greater emphasis should be placed on extrolite profiles and growth characteristics in species descriptions (Geiser *et al.*, 2007). They looked at multilocus sequencing of the genus to determine if it provides accurate species recognition. Their findings stated that COI is not suitable for black *Aspergilli* identification. The investigation suggested that partial β-tubulin and calmodulin are more promising markers for identification of *Aspergillus* (Geiser *et al.*, 2007).

Several experiments suggested that COI showed relatively low levels of variability to be useful for plant identification and barcoding (Kress *et al.*, 2005). Three criteria were set for the appropriate genetic loci; 1) significant species-level genetic variability and divergence, 2) an appropriately short sequence length so as to facilitate DNA extraction and amplification, and 3) the presence of conserved flanking sites for developing universal primers.

In an effort to find a correspondence between traditional species boundaries established by taxonomy and those inferred by DNA barcoding, Hebert *et al.* sequenced

DNA barcodes of 260 of the 667 bird species that breed in North America (Hebert *et al.*, 2004). They found that every single one of the 260 species had a different COI sequence. 130 species were represented by two or more specimens; in all of these species, COI sequences were either identical or were most similar to sequences of the same species. COI variations between species averaged 7.93% whereas variation within species averaged 0.43% (Hebert *et al.*, 2004). In four cases, there were deep intraspecific divergences, indicating possible new species. Three out of these four polytypic species are already split into two by some taxonomists (Hebert *et al.*, 2004). Theses results reinforced the view that COI is an important barcode region and strengthened the case for its use in DNA barcoding.

In another study, the red macroalgae employed DNA barcoding. Marine macroalgae is known to be extremely difficult to identify for numerous reasons. Although, the morphology and anatomy of the organism is very simple, rampant convergence, phenotypic plasticity, and alternation of heteromorphic generations contribute to its complex nature and difficulty of its proper identification (Saunders, 2005). COI was used extensively with fitting results. Novel primers were developed for red algae sequencing 50 COI 'barcode' regions and used to assess three known enigmatic and confusing species-complex questions involving: 1) *Mazzaella* species in the Northeast Pacific, 2) species of the genera *Dilsea* and *Neodilsea* in the Northeast Pacific, and 3) Asteromenia peltata from three oceans (Saunders, 2005). In all cases, the DNA barcode resolved the phylogeny accurately, identified confusing species identities, and turned up a variety of novel observations in need of further taxonomic investigation (Saunders, 2005).

Recently the efficacy of COI for DNA barcoding focused on the neotropical skipper butterfly *Astraptes fulgerator* at the Area Conservacion de Guanacaste (ACG) in north-western Costa Rica. This species was already known as a cryptic species, due to subtle morphological differences, as well as an unusually large variety of caterpillar food plants. However, several years would have been required for taxonomists to completely describe and delimit species within the complex (Hebert *et al.*, 2004). The group sequenced the COI gene of 484 specimens from the ACG. This sample included at least 20 individuals reared from each species of food plant, extremes and intermediates of adult and caterpillar color variation, and representatives from the three major ecosystems where *Astraptes fulgerator* is found (Hebert *et al.*, 2004). The findings concluded that *Astraptes fulgerator* consists of 10 different species in northwestern Costa Rica. Brower *et al.*, (2005) pointed out numerous flaws in the analysis, and, subsequently challenged these results. It was concluded that the original data could support no more than the possibility of three to seven cryptic taxa rather than ten cryptic species (Brower, 2005). This highlights that DNA barcoding analysis is dependent upon the choice of analytical methods used by the investigators, so the process of delimiting cryptic species using DNA barcodes can be as subjective as any other form of taxonomy.

Opposition

DNA barcoding has been met with spirited reaction from scientists, especially systematists, ranging from enthusiastic endorsement to outright opposition (Rubinoff *et al.*, 2006). For example, many scientists stress the fact that DNA barcoding does not provide reliable information above the species level; others indicate that it is inapplicable at the species level but may still have merit for higher-level groups (Whitworth *et al.*,

2007). Others resent what they see as a gross oversimplification of the science of taxonomy. Furthermore, some suggest that recently diverged species might not be distinguishable on the basis of their COI sequences (Kerr *et al.*, 2003). Recently, 23% of animal species were found to be polyphyletic if their mtDNA data are accurate, indicating that using mtDNA as a barcode to assign a species name to an animal will be ambiguous or erroneous 23% of the time (Funk *et al.*, 2003; Meyer *et al.*, 2005). Studies with insects suggest an equal or even greater error rate due to the frequent lack of correlation between the mitochondrial genome and the nuclear genome or the lack of a barcoding gap (Whitworth *et al.*, 2007; Hurst *et al.*, 2005; Wiemers *et al.*, 2007). Problems with mtDNA arising from male-killing microorganisms and cytoplasmic incompatibility-inducing symbionts are also particularly common among insects (Johnston *et al.*, 1996). Given that insects represent over 75% of all known organisms, this suggests that while mtDNA barcoding may work for vertebrates, it may not be effective for the majority of known organisms.

CHAPTER 3

DESIGN OF THE STUDY

GenBank Search

An initial search of all the completed COI genes for *Aspergillus* made available in GenBank were downloaded. The same was repeated for three *Penicillium* species, *Penicillium verrucosum*, *Penicillium viridiatum*, *Penicillium vulpinum*, and *Talaromyces trachysphermus*. The *Aspergilli* sequences were then scrutinized for multiple sequences within a species. *Aspergillus niger* and *Aspergillus tubingensis* both had abundant sequences available. Three samples each of the *Aspergillus niger* and *Aspergillus tubingensis* species were retained with the *Aspergillus* samples for a total of 19 downloaded genomic sequences from GenBank. The remaining sequences were not used for this study. The files were saved as FASTA files and exported to Geneious Pro© (Biomatters, Ltd. New Zealand).

Culturing Isolates

Five (5) isolates of *Aspergillus* and one (1) unknown isolate named "Vent Green" completed the number of subjects for which COI sequences were analyzed in the study. The *Aspergillus* isolates were obtained from the American Type Culture Collection (ATCC) and the Vent Green sample was acquired from the Texas Southern University Science Building ventilation system. These strains were cultured on malt extract and potato dextrose agar.

The malt and potato agar were prepared by adding 33.6 g and 39g respectively to 1.0 liter of purified water. Agar was heated to a boil and stirred for 1 minute to homogenize the mixture. The agar was then autoclaved at 121° C for 15 minutes. Once the agar cooled, it was plated in a sterile hood and left to solidify. Isolates were cultivated and then grown at 30°C for 4-7 days in an incubator.

DNA Isolation

The MO Bio UltraClean™ Microbial DNA Isolation Kit was used to isolate DNA from the fungal isolates. With each isolate, 1.8 ml of the fungal culture was added to a 2ml collection tube and centrifuged for 30 seconds at 10,000 x g at room temperature. The supernatant was decanted, and the tubes were spun once again for 30 seconds at 10,000 x g. Using a pipet tip, the supernatant was completely decanted leaving behind a pellet of microbial cells. The cell pellet was resuspended in a 300µl solution that contains a salt buffer that stabilizes and homogenously disperses the microbial cells prior to lysis. The solution was then vortexed and transferred to a tube of microbeads. To lyse the cells, 50µl of Solution MD1 containing sodium dodecyl sulfate (SDS) and other disruption agents are added. On a flat-bed vortex, the microbead tubes were secured and vortexed at maximum speed for 10 minutes. To separate the cell debris from the DNA, they are centrifuged at 10,000 x g for 30 seconds at room temperature. The supernatant was once again transferred to a clean 2 ml collection tube. 100 µl of Solution MD2, a reagent that removes contaminating non-DNA organic and inorganic material, was added to increase DNA purity for downstream DNA applications. It is then vortexed for 5 seconds, incubated at 4°C for 5 minutes, and centrifuged at 10,000 x g for 1 minute at room temperature. Avoiding the pellet containing contaminated matter, the supernatant

was transferred to a clean 2 ml collection tube. 900 µl of a highly concentrated salt solution, MD3, was added to the supernatant and vortexed for 5 seconds. 700µl of the solution was loaded onto a spin filter provided in the MO Bio kit and centrifuged for 30 seconds at 10,000 x g at room temperature. The flow through was then discarded. The remaining solution loaded onto the spin filter and centrifuged. The residual flow through was discarded. About 300 µl of solution MD4 was added to the spin filter tube and centrifuged at room temperature for 30 seconds at 10,000 x g. This binds DNA to the membrane and the ethanol-based solution of MD4 removed residues of salt left in MD3. The flow through was then discarded. The spin filter tube was centrifuged at room temperature for 1 minute at 10,000 x g to remove the ethanol wash. The spin filter was removed and placed into a clean 2 ml collection tube. About 50 µl of solution MD5 elution buffer was added to the center of the white filter membrane, promoting release of bound DNA. After a 30 second spin at 10,000 x g, the spin filter was discarded and genomic DNA was kept refrigerated.

Polymerase Chain Reaction

The extracted DNA was diluted to a concentration of 1:100 in preparation for downstream Polymerase Chain Reaction (PCR). PCR primers used followed Seifert *et al.* (2007). There were three species of the family Trichocomaceae with COI sequence designed primers (1), *Penicillium marneffei* and two *Aspergillus* species. The conserved regions of interest were rather AT-rich. One forward primer (PenF1, 5′-GACAAGAAAGGTGA-TTTTTATCTTC) (Seifert *et al.*, 2007) hybridizes at site 104. The reverse primer hybridizes at site 674 (reverse primer AspR1 5′-GGTAATGATAATAATAATACAGCTG) (Seifert *et al.*, 2007). PCR

amplifications of genomic DNA from five species of the phylum Ascomycota, namely: *A. clavatus* ATCC 10058, *A. foetidus* 10061, *A. giganteus* ATCC 10059, *A. tamarii* ATCC 1005, *A. terreus* ATCC 10020, and *Vent Green* Texas Southern University Science Building Ventilation System was done. The high success rate shows that the species of *Penicillium* used to design the forward primer, is phylogenetically related to *Aspergillus.*

Downstream PCR amplification was done utilizing the following parameters. PCR reactions were performed in 50-µl volumes containing 5.0 µl of PCR-grade water, 25.0 µl of Amplitaq Gold (Applied Biosystems), 5.0 µl of each primer (4 µM), and 5.0 µl of DNA extract. The cycling conditions had a hold step of 5 minutes at 95°C, 38 cycles of denaturing time of 15 seconds at 95°C, annealing for 15 seconds at 55°C, and 40 seconds at 70°C, followed by 10 minutes at 72°C.

Gel Electrophoresis

A portion of the amplicons was run on a 1.5% agarose gel at 60 volts for 45 minutes. To produce the agarose gel, 300 ml of Tris-Acetate-EDTA (TAE) was added to 4.5g of agarose gel. Heat was applied to the mixture to homogenize the substances. Once the gel cooled, 5.0 µl of ethidium bromide was added to 30 ml of the still liquefied gel. The gel was then poured into an electrophoresis tray containing combs to form the wells. After 15-20 minutes, the combs were removed and the gel was prepared for loading. About 5µl of the ϕX174 DNA/Hae III digest marker was used as the standard in lane one. Blue loading dye was added for visualization during the electrophoretic run. Once completed, the gels were placed in the Kodak 212 Gel Imaging System for analysis. Distinct bands around the 500-600 bp marker region indicated successful amplification. The samples were then prepared for DNA purification.

DNA Purification

The PCR product was purified in preparation for sequencing using the Wizard© PCR Preps DNA Purification System. About 100µl of Direct PCR Purification Buffer was dispensed into a tube. About 30-300 µl of the PCR product was added. After vortexing, 1 ml of resin was added, and again vortexed briefly to mix. About 1ml of resin was added and vortexed briefly 3 times over a 1-minute interval. The syringe barrel was attached to the minicolumn. The minicolumn /syringe barrel was inserted into the vacuum manifold. Resin/DNA mix was added to the syringe barrel. A vacuum was applied to pull the liquid through the minicolumn. When all the liquid passed through the minicolumn, the vacuum was released. 2 ml of 80% isopropanol was added to the syringe barrel. The vacuum was applied to pull the solution through the minicolumn. Resin was dried by continuing to apply the vacuum for 30 seconds. The syringe barrel was removed, and the minicolumn was transferred to a 1.5 ml microcentrifuge tube. The microcentrifuge tube was centrifuged at 10,000 x g for 2 minutes. The minicolumn was transferred to a clean 1.5 ml microcentrifuge tube. 50 µl of nuclease-free water was added to the minicolumn. The minicolumn was centrifuged at 10,000 x g for 20 seconds at room temperature. DNA was then stored at -20°C.

DNA Sequencing

SeqWright (Houston, Texas) provided sequencing for all PCR products. Sequencing reactions were performed in both directions by using the PCR primers PenF1 and AspR1. Sequencing products were run on an Applied Biosystems 3730 DNA analyzer.

SeqWright started with a mixture of template DNA, deoxynucleotides (dNTPs), DNA Polymerase, and primers PenF1 and AspR1. Limited quantities of dideoxynucleotides (ddNTPs) were also added. These 4 ddNTPs (ddATP, ddGTP, ddCTP, and ddTTP) were each fluoresced with a different color and lacked a 3' hydroxyl group (OH). In the absence of the 3'OH group, the ddNTPs terminated elongation of the DNA strand.

The DNA template was denatured with applied heat. It was prepared as a single strand. The primer annealed to its compliment on the template. As the temperature cooled, a double strand formed with the template and primer. DNA polymerase attached to the 3' end of the primer to "read" the template.

Chain elongation proceeded normally until, by chance, DNA polymerase inserted a ddNTP instead of the normal dNTP. Some DNA strands succeeded in adding several hundred nucleotides before insertion of the dideoxy version halted the process.

At the end of this incubation period, gel electrophoresis was used to separate the fragments by size. Each of the ddNTPs fluoresced a different color when illuminated by the laser and a scanner printed out the sequence.

Sequence Analysis

Using Geneious Pro© (Biomatters, Ltd. New Zealand), the sequences were assembled, edited, and aligned to find regions of similarity. Then, they were compared against the consensus identity for divergence. The phylogenetic relatedness was assessed and mapped using the bioinformatic program Phylogenetic Analysis Using Parsimony (PAUP©).

CHAPTER 4

RESULTS AND DISCUSSION

Gel Electrophoresis

In the study, the PCR products were amplified in a 1.5% agarose gel. The amplicons were present around the 550-600 bp position when compared to the marker. This is consistent with the expected length for COI.

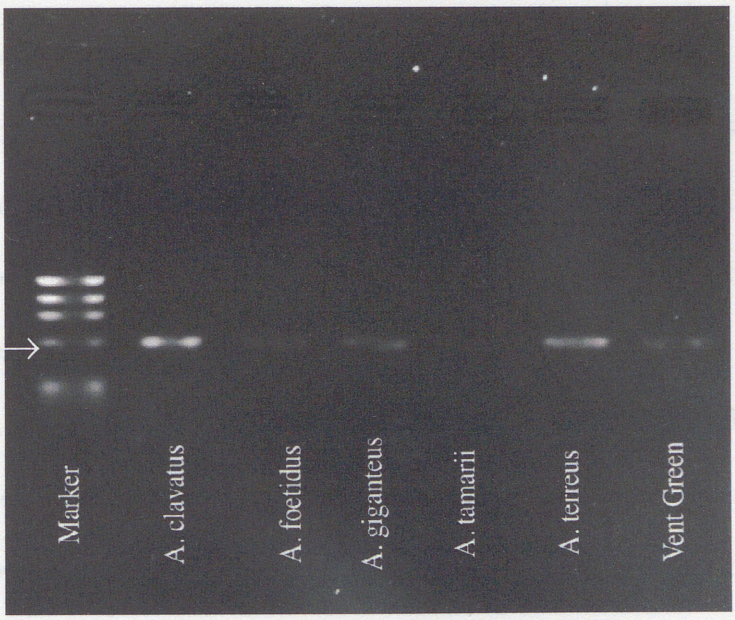


Figure 1: Electrophoresis Gel of Aspergillus fungal isolates

Chromatograms

DNA sequence chromatograms were assembled, with both forward and reverse primed sequences assembled as one consensus sequence. *Aspergillus clavatus* is represented in figure 2. Beginning at the 30th base, the graph showed distinct peaks with the exception at the 33rd and 559th base positions. The letter N indicated these ambiguities.

Aspergillus foetidus displayed in figure 3 illustrated distinct peaks in the sequence beginning at the 31st base position. *A. foetidus* had fewer ambiguities than that of *A. clavatus*.

The chromatogram of *Aspergillus giganteus* had a flawless reading frame starting at the 33rd base position. There were no uncertainties regarding the sequence of the isolate in figure 4.

In the representation of *Aspergillus tamarii* found in figure 5, the sequence became less clear beginning at the 257th base position. There were several ambiguities and the quality of the bases lessened. This was evident by the darkening of the highlighted bases.

Figure 6, was relatively devoid of uncertainties and contains high base quality. The last few bases were undetermined; however, *Aspergillus terreus* was a sound sample.

The chromatogram of the unknown isolate, Vent Green (fig. 7), illustrated an intact CO1 sequence. There were several bases that have low base quality, but overall, the sequence was useable for this study.

Each of the five sequences were subjected to Basic Local Alignment Search Tool (BLAST) search (Altschul *et al.*, 1990) to confirm sequence homology to archived sequences deposited in Genbank (http://www.ncbi.nlm.nih.gov). Of the five sequences, *A. foetidus* appeared to be questionable and unreliable, as the sequence resulted in 100 percent similarity with *A. niger*. *A. terreus* was 100 percent similar to archived *A. terreus*, and both were monophyletic in all the phylogenetic analysis.

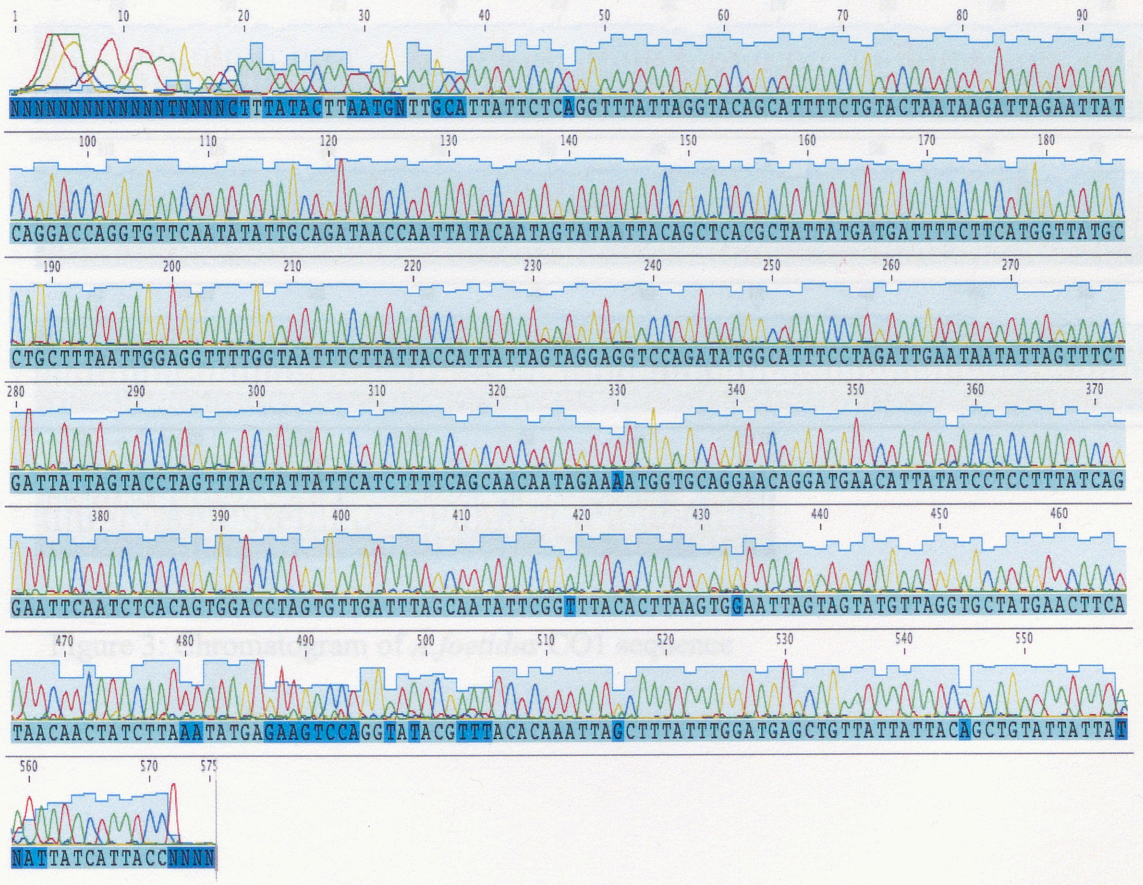


Figure 2: Chromatogram of *A. clavatus* CO1 sequence

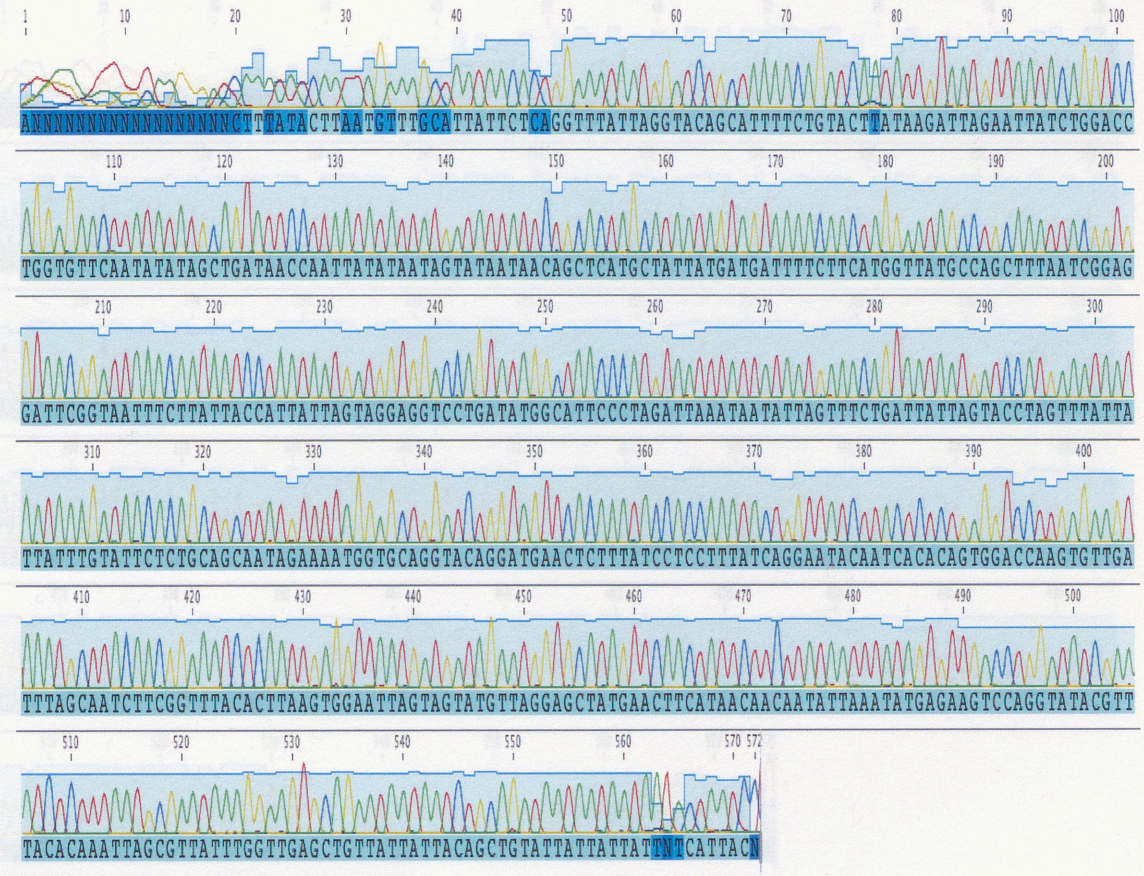


Figure 3: Chromatogram of *A.foetidus* CO1 sequence

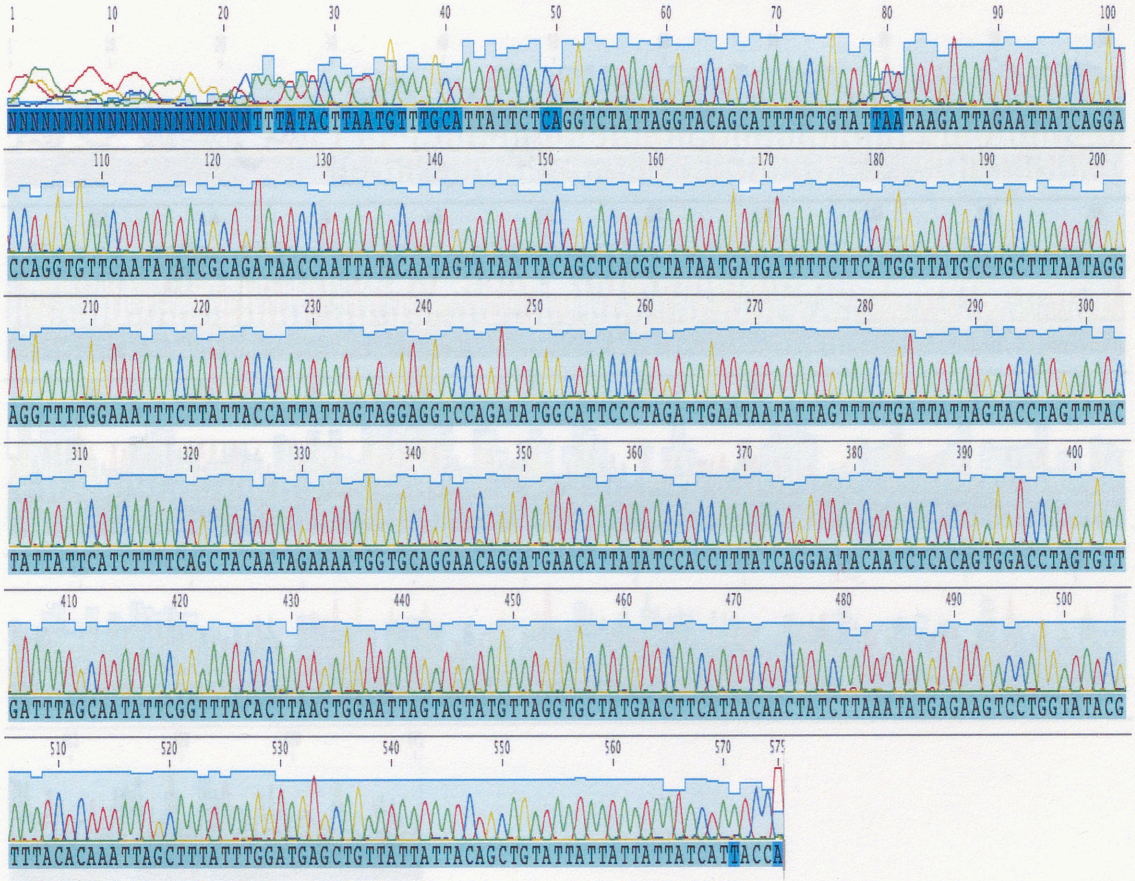


Figure 4: Chromatogram of *A. giganteus* CO1 sequence

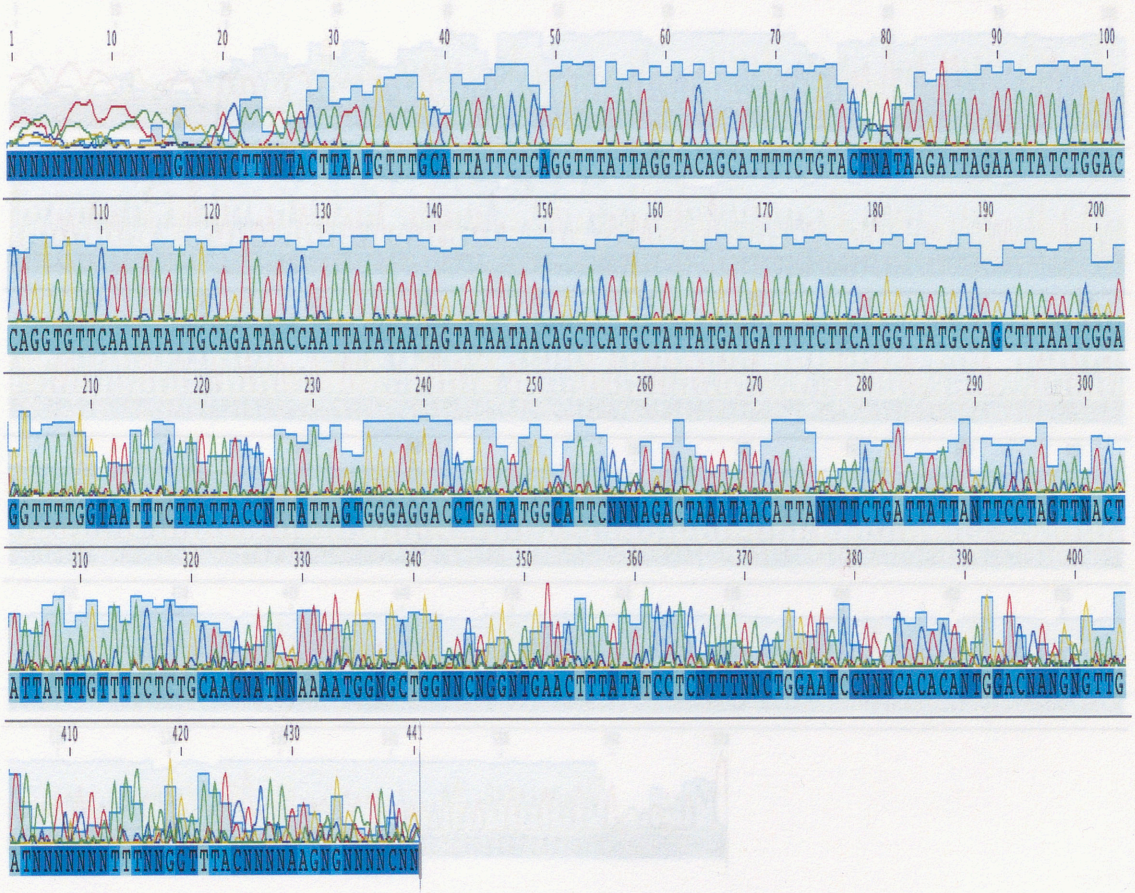


Figure 5: Chromatogram of *A. tamarii* CO1 sequence

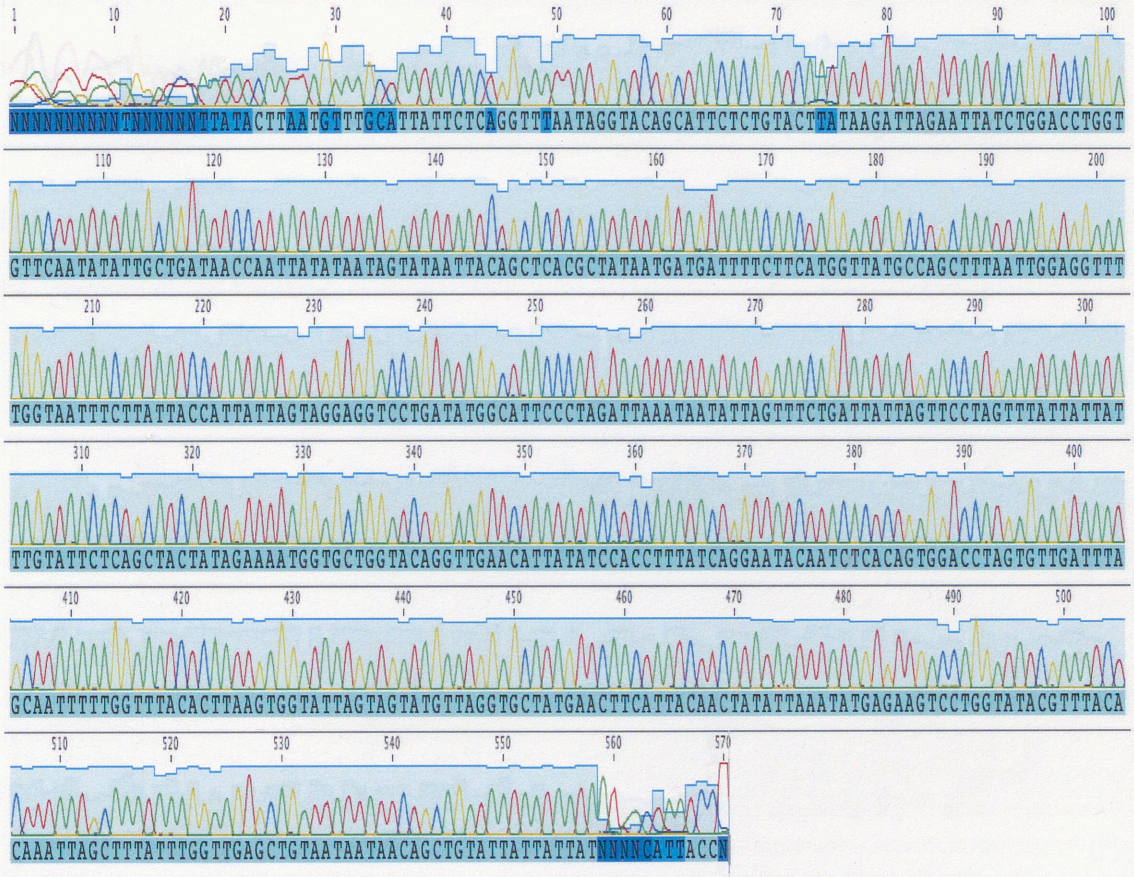


Figure 6: Chromatogram of *A. terreus* CO1 sequence

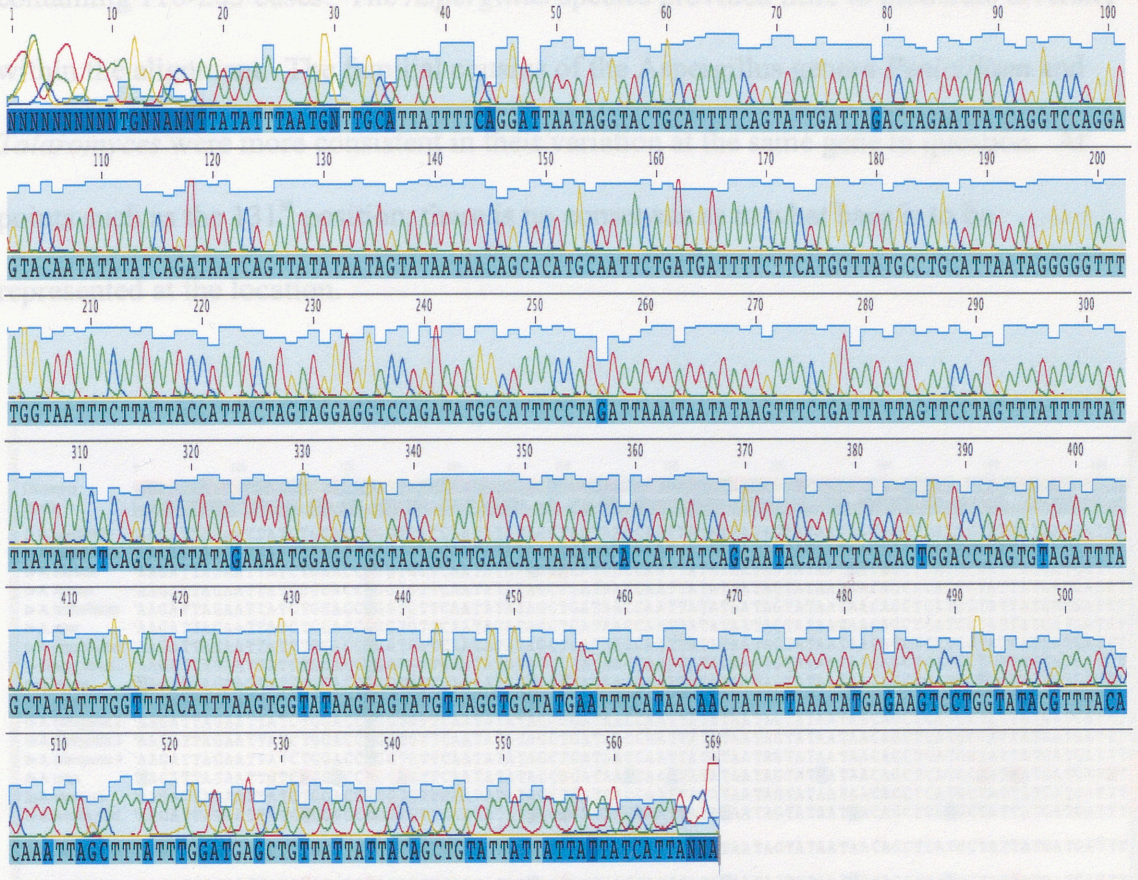


Figure 7: Chromatogram of unknown isolate “Vent Green"

Fungal Alignment

The base sequences from the chromatograms were aligned in Geneious Pro© and compared against the consensus identity for divergence. Figure 8 showed the region containing 110-203 bases. The *Aspergillus* species provided little to moderate diversity within the alignment. The familial cousins of the Aspergillus genera *Penicillium* and *Talaromyces* were more consistent in their variation at the same gene in question. At points such as the 131st position, there is no consensus as to what base is to be represented at the location.

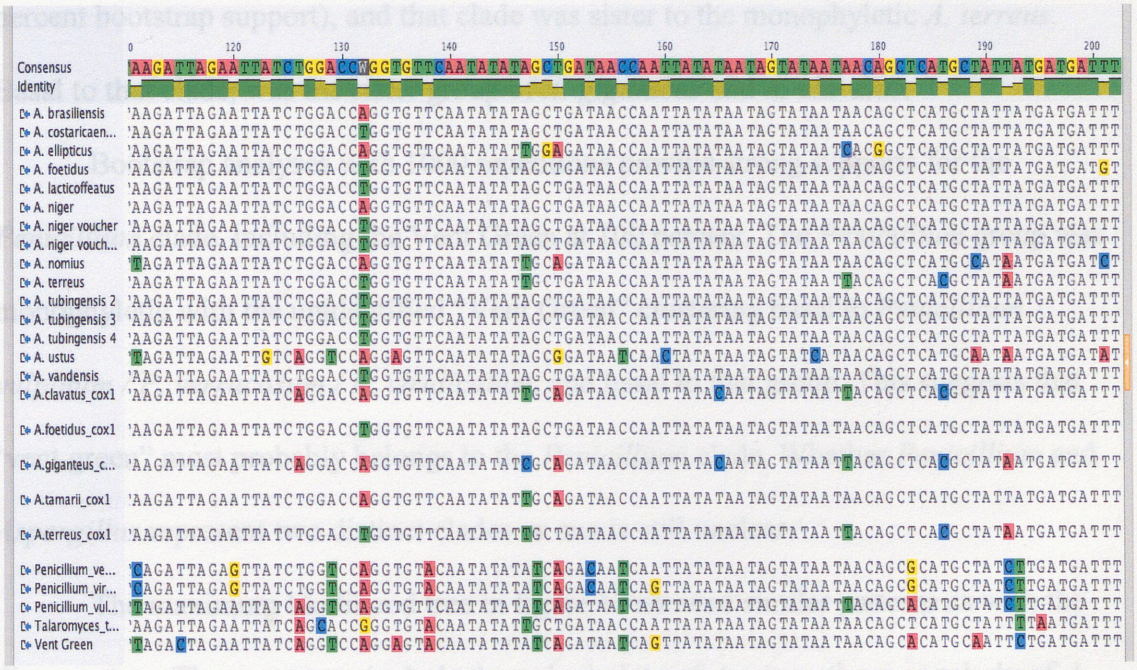


Figure 8: Fungal CO1 alignment

Phylogenetic Trees

The dataset was analyzed in three different optimality criteria using PAUP© 4.0 (Swofford 2002). All analyses were unrooted making no assumptions about the ancestry of the fungal samples. Parsimony analysis in figure 9 showed *A. tamarii* sister to *A. nomius*. That clade was basal to *A. terreus* clade. *A. clavatus* was sister to *A. giganteus*, and that clade was basal to the *A. terreus* clade. The lower region of the clade showed that the unknown isolate of "Vent Green" was clustered basal to the *Penicillium* genera. This is interesting in that it showed a relatedness of the unknown to *Penicillium*.

In the neighbor joining tree of figure 10, *A. tamarii* was sister to *A. nomius* (99 percent bootstrap support), and that clade was sister to the monophyletic *A. terreus*. Basal to that clade, was the sister group of *A. giganteus* and *A. clavatus*.

Bootstrap analyses with 1000 repetitions provide strong supports for the *Penicillium* clade, including the Vent Green at 100 percent. The *Penicillium* group was monophyletic with the unidentified "Vent Green" clustering basal to *Penicillium vulpinum / P. verrucosum / P. viridicatum*, and basal to that group. This suggests that 'vent green" most probably belongs to the *Penicillium* clade. Whether *Penicillium* and *Aspergillus* represent two distinct clades or one is still unclear.

The maximum likelihood tree in figure 12 showed similar patterns as with the distance tree. These patterns include the polyphylety of *A. niger*, the monophyly of *A. terreus* sequenced and *A. terreus* from Genbank, and the monophyly of *A. giganteus* and *A. clavatus*.

The three *Aspergillus niger* sequences obtained from Genbank formed a polyphyletic group, with one *A. niger* clustering with *A. foetidus*, and another *A. niger*

clustered with *A. vandensis/A. brasiliensis*. This ambiguous positions of the three *A. niger* from Genbank supports the notion that a significant amount of fungal information in Genbank is either erroneous, misclassification or contamination (Nillson *et al.* 2006; Bridge *et al.* 2003)

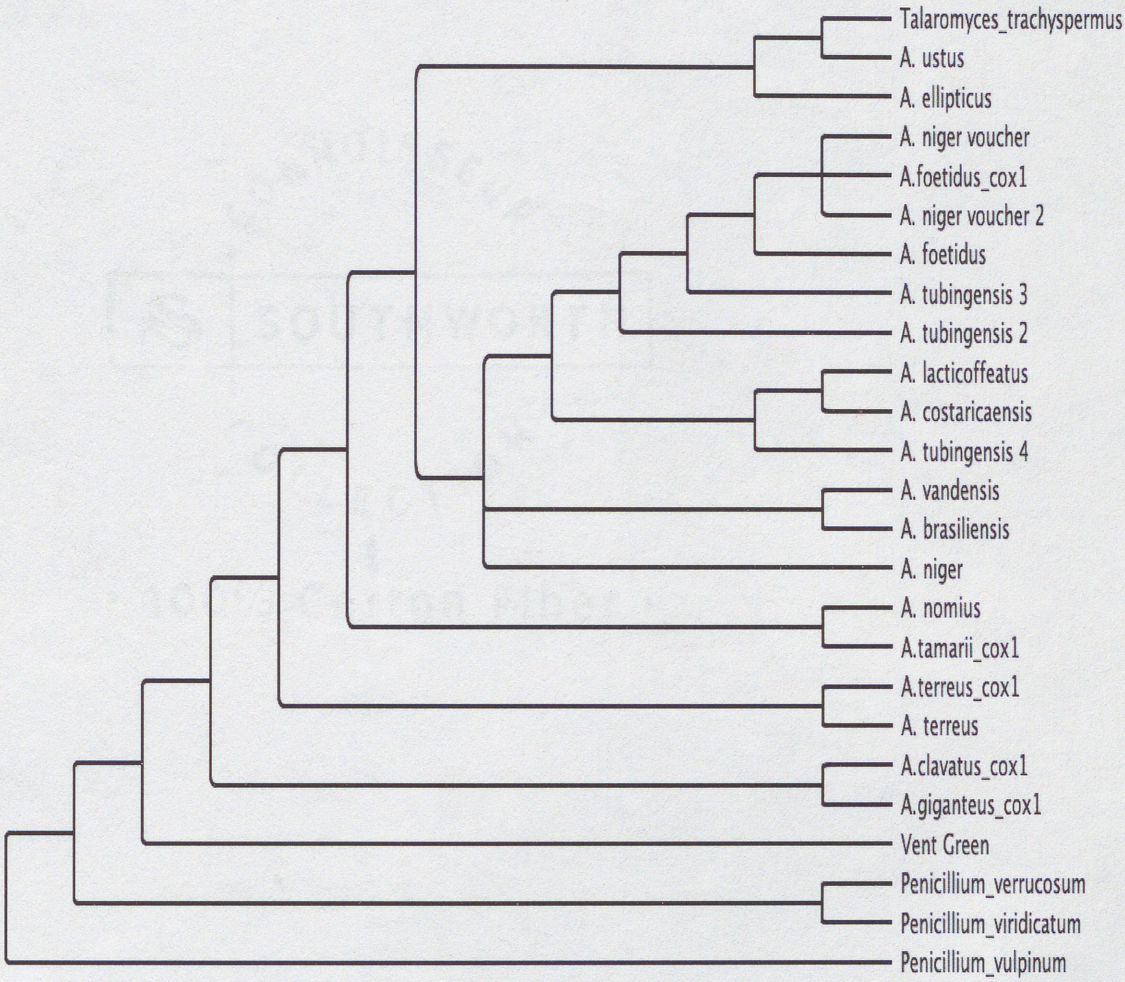


Figure 9: Maximum Parsimony Tree from fungal data set

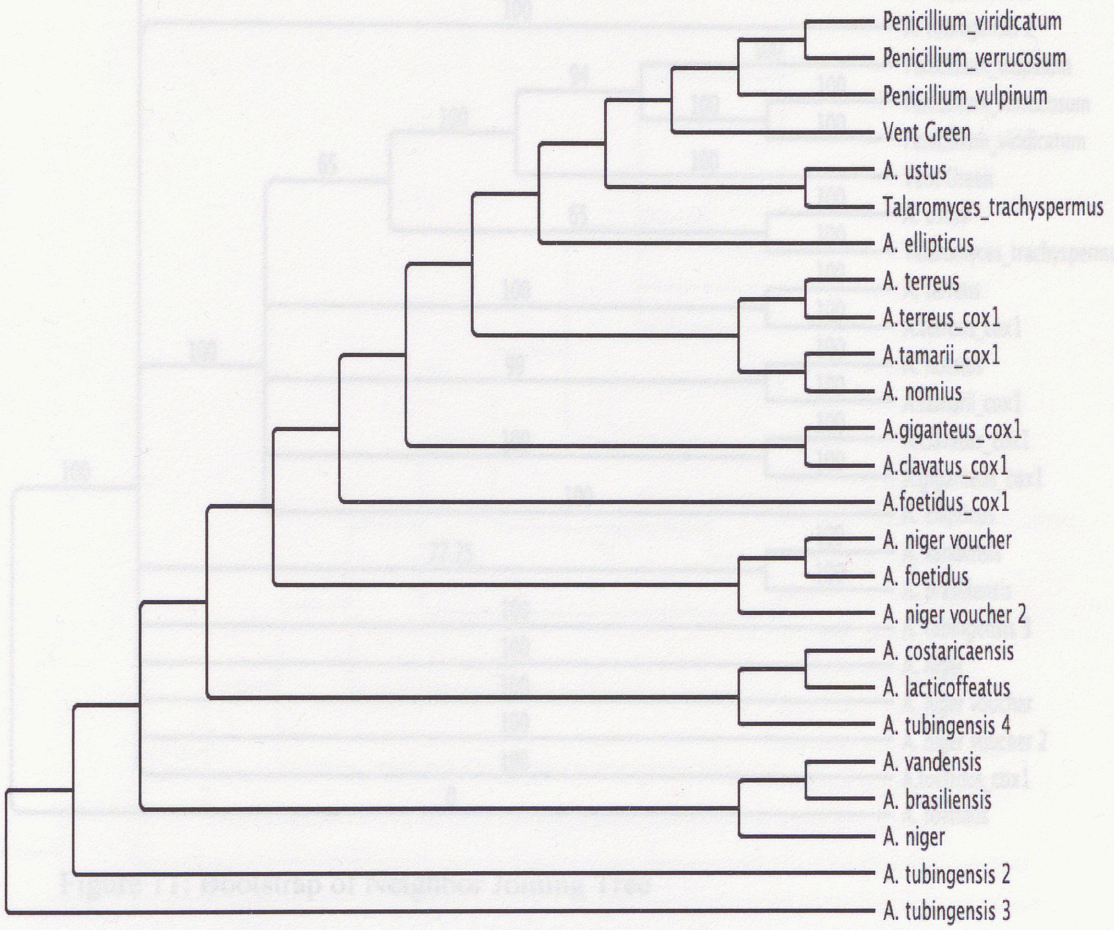


Figure 10: Neighbor Joining Tree from fungal data set using Jukes-Cantor model nreps=1000

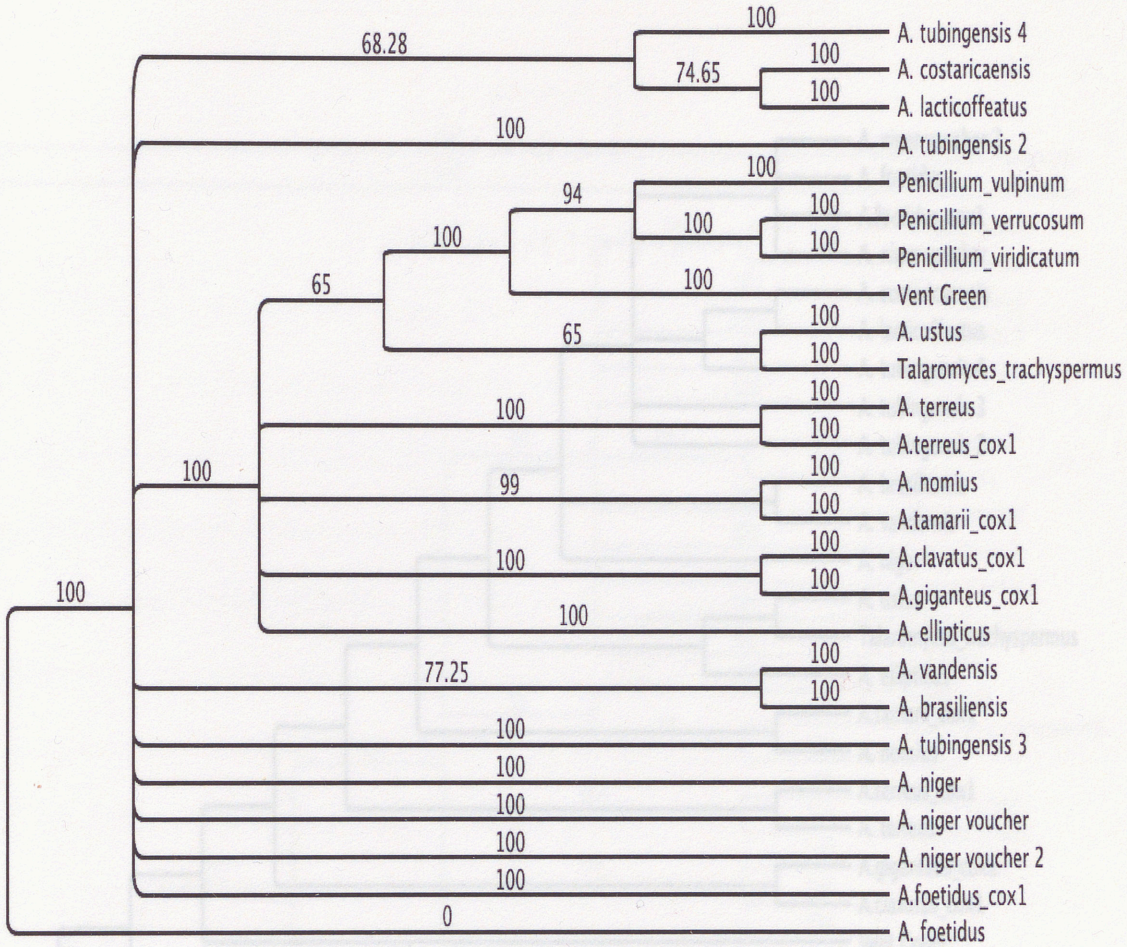


Figure 11: Bootstrap of Neighbor Joining Tree

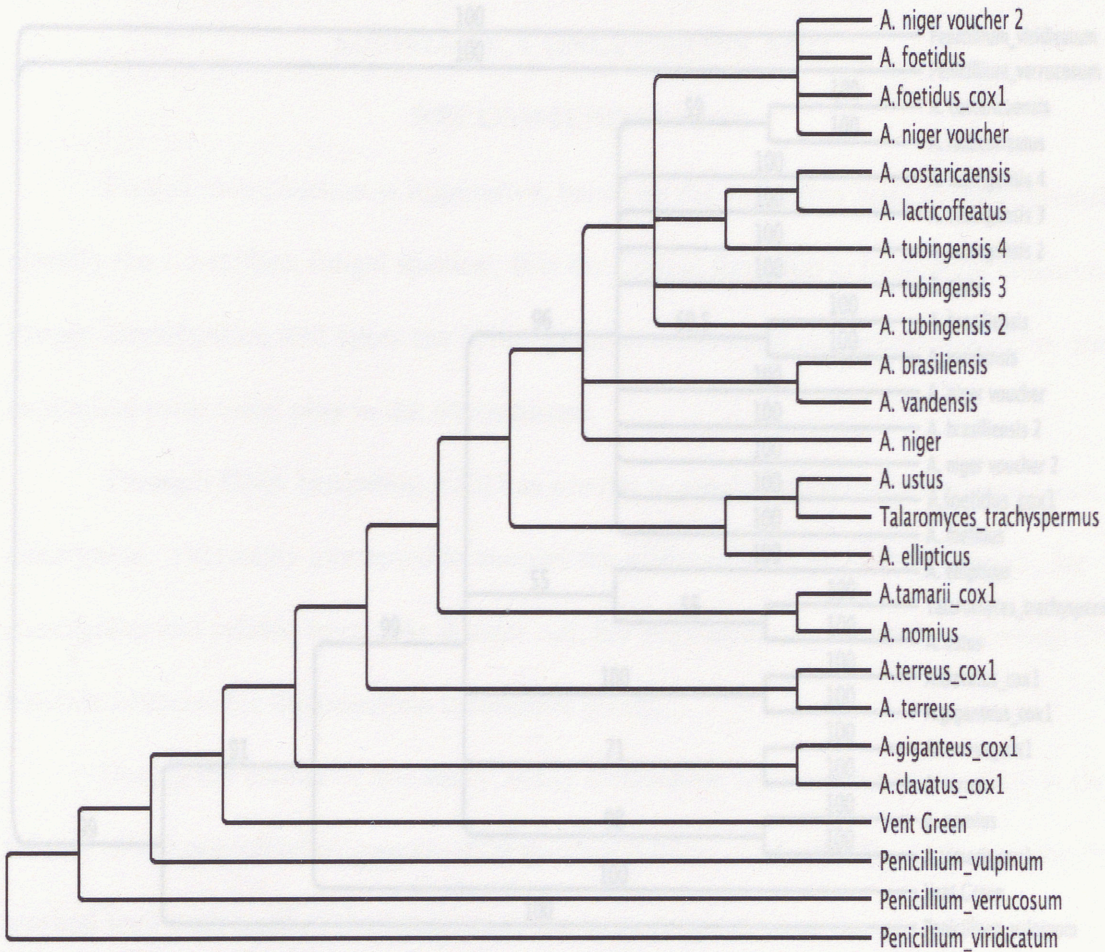


Figure 12: Maximum Likelihood Tree from the fungal data set using Jukes-Cantor Model

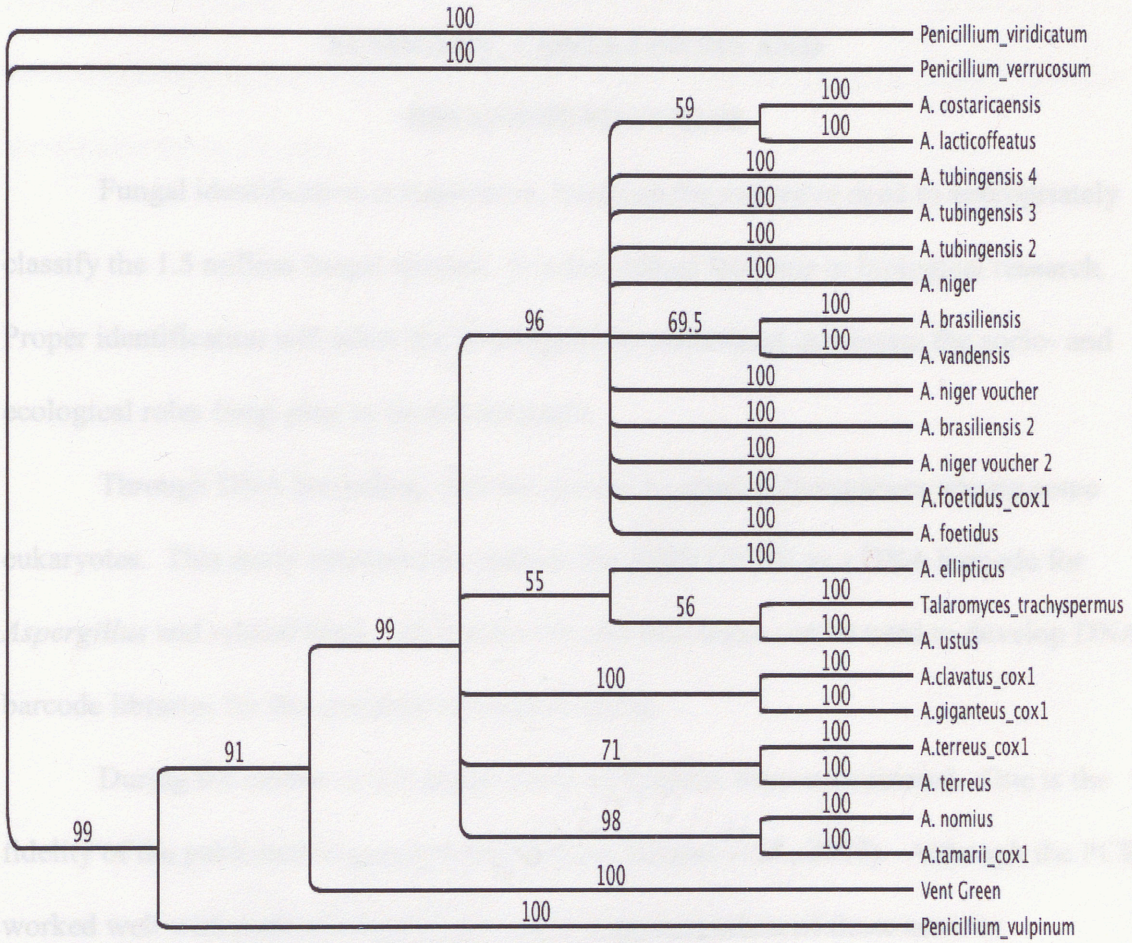


Figure 13: Bootstrap of Maximum Likelihood Tree
nreps=1000

CHAPTER 5

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Fungal identification is imperative, based on the extensive need to appropriately classify the 1.5 million fungal species. It is the critical first step in biological research. Proper identification will allow the investigator to understand and assess the socio- and ecological roles fungi play in the environment.

Through DNA barcoding, COI has proven to resolve divergences among some eukaryotes. This study attempted to analyze the utility of COI as a DNA barcode for *Aspergillus* and related fungi. As it turns out, the sequences can be used to develop DNA barcode libraries for the complete taxonomic group.

During the course of this study, many difficulties were encountered. One is the fidelity of the published oligonucleotide primers (Geiser *et al.*, 2007). Although the PCR worked well with most of the taxa assayed, a large proportion of those sent for sequencing came back with undistinguishable chromatograms. This observation also applies to fungal sequencing efforts based on another gene used in the lab– the Internally Transcribed Spacer (ITS) region (unpublished). Several sources of sequencing difficulty may include: 1) a less stringent and less effective method of genomic isolation, 2) presence if inhibitory nucleases in fungal genomic samples that inhibit the amplification of target genes, or 3) non-optimal PCR parameters. It is recommended that new sets of primers should be designed that should work on a broader taxonomic range.

Despite reports of the presence of introns within COI among fungi (Seifert *et al.*, 2007), this was not observed in the five samples that we successfully sequenced. My preliminary data suggests that COI can potentially be useful for identification and phylogenetic reconstructions of fungi. However, a far larger dataset for sampling is needed to ascertain this observation. Moreover, other loci of the mitochondrial genome should also be investigated.

REFERENCES

1. Altschul, S. F., W. gish, W. Miller, E. W. Myers, D. J. Lipman. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. 215: 403-410.

2. Bennett JW. "An Overview of the Genus *Aspergillus*". (2010). *Aspergillus*: *Molecular Biology and Genomics*. Caister Academic Press.

3. Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. (2005). Defining operational taxonomic units using DNA barcode data. *Phil Trans R Soc B*, 360:1935-1943.

4. Bridge P.D., P.J. Roberts, B.M. Spooner, G. Panchal. (2003). On the unreliability of published DNA sequences. *New Phytol* 160: 43–48.

5. Brower AVZ: Problems with DNA barcodes for species delimitation. (2005). 'Ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Systematics and Biodiversity* 4(2): 127–32.

6. Funk DJ, Omland KE: Species-level paraphyly and polyphyly. (2003). frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Syst*, 34: 397–423.

7. Geiser DM, Klich MA, Frisvad JC, Peterson SW, Varga J, Samson RA. (2007) The current status of species recognition and identification in Aspergillus. *Stud Mycol*, 59:1-10.

8. Hajibabaei M, Singer GA, Clare EL, Hebert PDN. (2007). Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biol*, 5:24.

9. Hawksworth DL. (2004). Fungal diversity and its implications for genetic resource collections. *Studies in Mycology*, 50, 9–18.

10. Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research*, 105: 1422–1432.

11. Hawksworth DL. (1991). The fungal dimension of biodiversity, magnitude, significance, and conservation. *Mycol Res*, 95:641-655.

12. Hebert PD, Stoeckle MY, Zemlak TS, Francis CM. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, 2(10):E312.

13. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proc Natl Acad Sci USA*, 101(41):14812-814817.

14. Hebert PDN, Ratnasingham S, deWaard JR. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci*, 270(Suppl 1):S96-S99.

15. Hillis, D.M., C. Moritz, and B. K. Mable. (1996). Molecualr Systematis, 2nd ed. Sunderland, MA: Sinauer Associates, Inc.

16. Hurst GD, Jiggins FM. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. Biol. Sci*, 272(1572): 1525–34.

17. Johnstone RA, Hurst GDD. (1996). Maternally inherited male-killing microorganisms may confound interpretation of mitochondrial DNA variability. *Biol. J. Linnaean Soc.* 58: 453–70.

18. Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CN, Hebert PDN. (2007). Comprehensive DNA barcode coverage of North American birds: *Molecular Ecology Notes*, 7(4):535-543.

19. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. (2005). Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA*, 102(23):8369-8374.

20. Kubicek CP, Komon-Zelazowska M, Druzhinina IS. (2008). Fungal genus Hypocrea/Trichoderma: from barcodes to biodiversity. *J Zhejiang Univ Sci B* 2008, 9(10):753-63.

21. Meyer CP, Paulay G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol*, 3(12): e422.

22. Min XJ, Hickey DA. (2007). Assessing the effect of varying sequence length on DNA barcoding of fungi. *Mol Ecol Notes*, 7(3):365-373.

23. Nillson, R. H., R. Ryberg, E. Kristiansson, K. Abarenkov, K.-H. Larrson, U. Koljalg. (2006). Taxonomic reliability of DNA sequence in public sequence databases. *PLoS ONE* 1: e59.

24. Rubinoff D, Cameron S, Will K. (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification: *J. Hered.*, 97 (6): 581–94.

25. Santamaria M, Vicario S, Pappadà G, Scioscia G, Scazzpcchio C, Saccone, C. (2009). Towards barcode markers in Fungi: an intron map of Ascomycota mitochondria. *BMC Bioinformatics,* 10(Suppl 6):S 15.

26. Saunders GW. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos Trans R Soc Lond B Biol Sci*, 360:1879–1888.

27. Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. (2005). Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philos Trans R Soc London B*, 360:1805–1811.

28. Seifert, K. A. (2009). Progress towards DNA barcoding of fungi. *Molecular Ecology Resource* 9 (Suppl.1), 83-89.

29. Seifert KA, Samson RA, Dewaard JR, Houbraken J, Lévesque CA, Moncalvo JM, Louis-Seize G, Hebert PD. (2007). Prospects for fungus identification using CO1 DNA barcodes, with Penicillium as a test case. *Proc Natl Acad Sci USA*, 104(10):3901-6.

30. Smith MA, Fisher BL, Hebert PDN. (2005). DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos Trans R Soc Lond B Biol Sci*, 360:1825-1834.

31. Swofford, D. L. (2002). PAUP*: Phylogenetic analysis using parsimony (and other methods). Sinauer Associates.

32. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005). DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci,* 360:1847-1857.

33. Whitworth TL, Dawson RD, Magalon H, Baudry E (2007). DNA barcoding cannot reliably identify species of the blowfly genus Protocalliphora (Diptera: Calliphoridae). *Proc. R. Soc. B*, 274:1731–1739.

34. Wiemers M, Fiedler K. (2007). Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool*, 4: 8.