Biochemical Markers for the Detection and Classification of *Aspergillus*

Thesis presented in candidature

for the degree *Philosophiae Doctor*

Norman J. Glassbrook, M.S.

Cardiff School of Biosciences

Cardiff University

Cardiff

United Kingdom

September 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

Many people have contributed to the development and completion of the work presented here by providing encouragement, guidance, technical support and resources. I thank Gary Payne and Becky Boston of NC State for opening their laboratories to me and helping me become a part of the NC State University system. Many thanks to Nigel Deighton and Gary Payne of NC State for serving as local advisors and providing the technical guidance and resources I needed to complete the study. I thank J. Richard Dickinson for encouraging me to pursue a research doctorate through Cardiff University and serving as an advisor and supervisor for my research. I thank the research groups of Nigel Deighton, Gary Payne, Becky Boston, Margaret Daub, Ralph Dean and Charles Opperman at NC State for technical guidance and laboratory resources. I thank Fred Jaeger of Reichhold for encouragement and technical support before I obtained laboratory space at NC State. Last, but certainly not least, I thank my wife, Sally Van Wert, and two sons, Jason and Keith Glassbrook, for their encouragement, support and patience during the entire process.

Abstract

The genus *Aspergillus* includes a diverse group of filamentous fungi that are widely distributed in nature, commonly found in soil. The Aspergilli include species that can be beneficial or detrimental to humans, so detection and accurate identification of these organisms can be very important. Morphology and genetic sequence analysis are well established methods for classifying and identifying fungi, but morphology remains a widely used technique that generally works well for Aspergilli. However, some organisms may be misidentified due to atypical morphology and some hidden (cryptic) species may not be recognized as different from named species based on readily observable traits. In this study, reference strains of different *Aspergillus* species, *Penicillium chrysogenum*, *Candida albicans*, and *Cryptococcus neoformans* were characterized using LC/MS and GC/MS biochemical profiling techniques in order to find specific small molecules, peptides or biochemical profiles that can be used in addition to established methods to detect and classify Aspergilli to the species level. Subsequently, analytical methods developed for characterizing the reference strains were applied, along with morphology and PCR, to characterize and identify several laboratory and field isolates. Some unique compounds and biochemical patterns did emerge from small molecule profiling that could be used for classifying Aspergilli, but protein profiling by LC/MS/MS was a much more effective approach. Tandem mass spectra from LC/MS/MS of tryptic peptides from fungal proteins were searched against protein databases and matched to theoretical spectra derived from those databases. Many of the amino acid sequences detected were taxonomically diagnostic for classifying *Aspergillus* species. Protein profiling also provided a great deal of additional biochemical information on the test organisms by identifying the predominant enzymes and structural proteins present under different experimental conditions and may find broader application for identifying and studying other organisms.

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Figures

List of Figures (continued)

Chapter 3: Results of GC/MS Analyses

Chapter 4: Results of LC/MS Analyses

List of Figures (continued)

Chapter 4: Results of LC/MS Analyses

Chapter 5: Case Studies

List of Figures (continued)

Chapter 5: Case Studies

List of Figures (continued)

Chapter 5: Case Studies

List of Abbreviations

BCP	- biochemical profiling
CxP	- a mixture of alkyl pyrrolidones
EDTA	- ethylenediaminetetracetic acid
EI	- electron impact ionization
ESI	- electrospray ionisation
GAPDH	- glyceraldehyde-3-phosphate dehydrogenase
GC	- gas chromatography or gas chromatograph
HED	- hydroxyethyldisulfide
LC	- liquid chromatography or liquid chromatograph
MS	- mass spectrometry
MS/MS	- tandem mass spectrometry
NCBI	- National Center for Biotechnology Information
NIST	- National Institute of Standards and Technology
SDS	- sodium dodecylsulfate

Chapter 1: Biochemical Markers for Detection and Identification of *Aspergillus* Species - Overview

Introduction

The genus *Aspergillus* includes a diverse group of filamentous fungi that are widely distributed in nature, commonly found in soil. The Aspergilli include species that can be beneficial or detrimental to humans, so accurate identification of these fungi can be extremely important. As with other fungi, Aspergilli historically have been classified and identified based on morphology along with supplementary physiological, biochemical, life cycle and ecological characteristics (Ainsworth *et al.* 1965, Raper and Fennell 1965). Classification and identification based on morphology is still widely used and generally works well for Aspergilli, but some organisms may be misidentified due to atypical morphology and some hidden (cryptic) species may not recognized as different from named species based on readily observable traits. Increasingly, molecular techniques are being applied to identify unknown fungi and re-evaluate morphology-based classifications (Klich 2006, Sampson *et al.* 2006, Geiser *et al.* 2007). In this study, a number of *Aspergillus* reference species were characterized using LC/MS and GC/MS biochemical profiling techniques in order to find specific small molecules, peptides or biochemical profiles that might be used in addition to morphological and molecular techniques for the detection and identification of Aspergilli.

Overview of the Aspergilli

The genus *Aspergillus* includes a diverse group of filamentous fungi that are widely distributed in nature, commonly found in soil. They have been studied extensively due to their importance in fermentation, food safety, plant pathology, and medicine. Although most species are generally considered saprophytic, some are capable of causing disease in insects, plants, and animals (Gugnani 2003, Hedayati *et al.* 2007). This study focused on *Aspergillus flavus*, a species that can impact human and animal health in two distinctly different ways – as a producer of mycotoxins or as an opportunistic pathogen.

A number of *Aspergillus* species can contaminate plant products during production, processing, and storage. Aspergillus is one of the major genera of mycotoxigenic fungi (*Aspergillus*, *Penicillium*, *Fusarium*) and can produce an array of mycotoxins including aflatoxins, fumonisins, ochratoxins, tricothecenes, ergot alkaloids, sterigmatocystin, gliotoxin, cyclopiazonic acid, citrinin, and patulin (CAST 2002).

Aspergillus flavus is one of five *Aspergillus* species - *A. flavus, A. parasiticus, A. nomius, A. pseudotamarii, and A. bombycis* - that produces aflatoxins, fungal metabolites that are toxic and carcinogenic to humans and animals (CAST 2002, Scheidegger and Payne 2003). Closely related domesticated species, *A. oryzae* and *A. sojae*, are used in the fermentation industry and are non-aflatoxigenic. The aflatoxigenic species are frequently found in agricultural products such as maize, cotton and peanuts. Allowable concentrations for aflatoxins are regulated and commodities that exceed limits must be decontaminated or destroyed. The U.S. regulatory limits for aflatoxin are set at 0.5 ppb for milk, 20 ppb for food intended for human consumption, and 20-300 ppb for animal feed; regulatory limits in the E.U. are set much lower at 0.05 ppb for milk and 2-15 ppb for spices and food (Murphy *et al.* 2006).

A. flavus is also one of several *Aspergillus* species that causes allergy or infection in humans. Common pathogenic species are *A. fumigatus*, *A. flavus*, *A. niger*, and *A. terreus*, with *A. fumigatus* the most prevalent. *Aspergillus* conidia, particularly those of *A. fumigatus*, are small enough to travel through the air and into the lungs. Normal, healthy individuals rarely develop aspergillosis because the immune system keeps the fungus in check. However, people weakened by other diseases, or individuals under immunosuppressive therapy for organ transplants or treatment of autoimmune diseases, are at greatly increased risk of developing aspergillosis (Denning 1998, Latge 1999, Yaguchi *et al.* 2007).

Historical Perspective

Fungi have a very long lineage based on a fossil record that extends back beyond 500 million years and there are many references to fungi throughout human history (Ainsworth 1965, Lutzoni *et al.* 2004, Blackwell *et al.* 2009 [tolweb]). The first scientific descriptions of Aspergilli appeared after the introduction of the microscope and are attributed to Pietro Antonio Micheli in "Nova Genera Plantarum" published in 1729 (Ainsworth 1965, Raper and Fennel 1965, Gugnani 2003). Using microscopy, Micheli was able to distinguish stalks and spore heads as well as the spore chains that radiated out from a central structure. He named these molds *Aspergillus* because these structures resembled an aspergillium (holy water sprinkler). Much of the early literature is based on this type of microscopic study of specimens on their natural substrates.

The study of Aspergilli grew through the 1800's with the development of laboratory culture techniques by De Bary and others as well as the increasing use of these fungi in industrial fermentations. A series of monographs on the genus have been written

over the years to summarize the collective information available for Aspergilli including Wehmer in 1901, Thom and Church in 1926, Thom and Raper in 1945, and Raper and Fennel in 1965 (Raper and Fennel 1965, Gugnani 2003). Raper and Fennel recognized 132 *Aspergillus* species; currently the total number of recognized species is on the order of 250 (Geiser *et al.* 2007)

Taxonomy and Species Concept in Aspergilli

A standard definition of species is "a fundamental taxonomic category ranking below a genus and consisting of a group of closely related individuals that can interbreed freely and produce fertile offspring" (Oxford Dictionary of Biochemistry and Molecular Biology, Smith 2000). The recognition of biological species based on mating tests is not possible with many Aspergilli because only about a third of *Aspergillus* species have a known sexual stage or teleomorph (Geiser 2008). Consequently, morphological and other phenotypic traits of the asexual or anamorphic forms, as described by Raper and Fennel (1965), have served as the basis of classifying many of the Aspergilli to the species level. This traditional approach to *Aspergillus* taxonomy has worked well and is still central to the defining the genus but the classifications have been refined and revised based on molecular techniques that derive phylogenetic relationships based on comparisons of DNA sequences (Samson *et al.* 2006, Geiser *et al.* 2007). The current taxonomic classification of the genus Aspergillus is summarized in Table 1.1.

Raper and Fennel subdivided the genus into 18 groups of one to several species based on probable relationships deduced from shared characteristics. The group concept was useful for classification, but the term had no formal taxonomic standing under the International Code of Botanical Nomenclature (ICBN) that governs fungal

Rank	Taxon	Classification Features
Domain	Eukaryota	organisms with nuclear membranes
Kingdom	Fungi	osmotrophic heterotrophs with chitinous cell walls, generally filamentous
Subkingdom	Dikarya	unicellular or filamentous, often dikaryotic (binucleate) lacking flagella
Phylum	Ascomycota	ascospores (meiospores) inside sac-like cases (asci)
Subphylum	Pezizomycotina	predominantly filamentous, septate
Class	Eurotiomycetes	diverse ascoma types, molecular phylogeny
Subclass	Eurotiomycetidae	mostly enclosed ascomata and prototunicate asci contains most fungi previously in Plectomycetes
Order	Eurotiales	cleistothecial ascomata, classically green and blue molds, molecular phylogeny
Family	Trichocomaceae (*syn.* Eurotiaceae)	cleistothecial characteristics, molecular phylogeny selected anamorphic genera include *Aspergillus*, *Paecilomyces*, *Penicillium*
Genus	*Aspergillus*	aspergillum-like asexual reproductive structures, molecular phylogeny

Table 1.1. Current taxonomic classification of the genus *Aspergillus*. Table compiled from Sugiyama 1998, Guarro *et al.* 1999, Lumbsch 2000, Gugnani 2003, Lutzoni *et al.* 2004, Geiser *et al.* 2006, Spatafora 2006, Hibbett *et al.* 2007, Stchigel and Guarro 2007, Geiser 2008, Humber 2008.

nomenclature (Samson 1992, Guarro 1999). Consequently, the groups were revised and given taxonomic status as sections and six subgenera were added (Gams *et al.* 1985, Samson 1992, Geiser *et al.* 2007). Peterson, in 2000, proposed 15 sections in three subgenera based on rDNA sequences. In 2005, Frisvad *et al.* suggested adding a new section, *Ochracerosei* to accommodate *A. ochraceoroseus* and a new species *A. rambellii.* In 2008, Peterson revised his infrageneric taxa to 17 sections

organized into 5 subgenera based on an expanded phylogenetic analysis of four loci – beta tubulin, calmodulin, internal transcribed spacer (ITS) and large subunit rDNA, and RNA polymerase II. This classification scheme is presented in Table 1.2.

Subgenus *Aspergillus*
- Section *Aspergillus*
- Section *Restricti*

Subgenus *Circumdati*
- Section *Candidi*
- Section *Circumdati*
- Section *Cremei*
- Section *Flavi*
- Section *Flavipedes*
- Section *Nigri*
- Section *Terrei*

Subgenus *Fumigati*
- Section *Cervini*
- Section *Clavati*
- Section *Fumigati*

Subgenus *Nidulantes*
- Section *Nidulantes*
- Section *Ochraceorosei*
- Section *Sparsi*
- Section *Usti*

Subgenus *Ornati*
- Section *Ornati*

Table 1.2. Nomenclature of the infrageneric taxa of the genus *Aspergillus* according to Peterson (2008). Classification assessment is based on multilocus testing of approximately 460 *Aspergillus* isolates.

Naming Conventions for Aspergilli

For an Aspergillus species to be formally recognized by taxonomists, it must be described appropriately and assigned a Latin binomial name in accordance with the bionomenclature rules of the ICBN (Samson 1992, Guarro 1999). Because Aspergilli have been observed and described based in teleomorph and/or anamorph forms, often separately and without knowledge of the relationship between the two forms, there is a history of dual nomenclature, with each form having its own genus name. Under the ICBN, if the teleomorph form is known, the teleomorph name takes precedence, but the use of the anamorph name is permitted, particularly when the organism is typically encountered and more widely known in the anamorph state (Cline 2005, Pitt and Samson 2007). The test organisms of the Aspergilli in this study are referred to as *Aspergillus*, the genus name of the asexual stage, throughout this text.

Current Methods for Detection and Identification

Although numerous immunological, molecular, and biochemical methods are now available, the morphological approach to identification of Aspergilli, based on their distinctive phenotypic characteristics in culture, remains a widely used technique (Samson *et al.* 2006, Klich 2006, Denning 1998). In fact, a 2003 survey of laboratories performing mycological testing reported that 89% use culture, 16% use serology, and fewer than 5% use molecular tests (McKlenny 2005).

Methods for detection, characterization, and identification of fungi, including *Aspergillus*, have been extensively reviewed elsewhere (Frisvad *et al.* 1998, Guarro *et al.* 1999, Raper and Fennel 1965, Rath 2001, Samson *et al.* 2006, Varga *et al.* 2004, Peterson 2008) and will only be summarized here. Microscopy and culture

were used in this study along with mycotoxin analysis to confirm the identity of species and strains, when cultured organisms produced asexual reproductive structures and presented species-typical morphology. Some reference strains produced only mycelia and lacked distinctive colours or other features that would aid identification, and so the identity could not be readily confirmed by culture phenotype.

Culture and Morphology. The genus *Aspergillus* is characterized by a distinctive asexual reproductive structure, the aspergillum, that resembles an aspergill brush or sprinkler, from which it gets its name. The aspergillum consists of a stipe terminating in a spore head or vesicle, on which spore forming cells and spores are borne, see Figure 1.1. Another feature characteristic of Aspergilli is the presence of foot cells, enlarged, thick-walled hyphal cells from which the stipe arises (Mackenzie 1988, Ratna 1973). Hyphae are hyaline (clear, glassy) and septate with acute branching (<90° angle) (Mackenzie 1988, McKlenny 2005).

Aspergilli are easily cultured on common laboratory media such as potato dextrose agar (PDA) and variants of Czapek (CZ) and malt extract agars (MEA). Many of the species, if grown on standardized media for specified lengths of time, can be classified macroscopically based on growth habit and colour and microscopically based on aspergillum and spore characteristics. Diagrams of typical condiophores and the general appearance of an *A. flavus* conidiophore is presented in Figure 1.2.

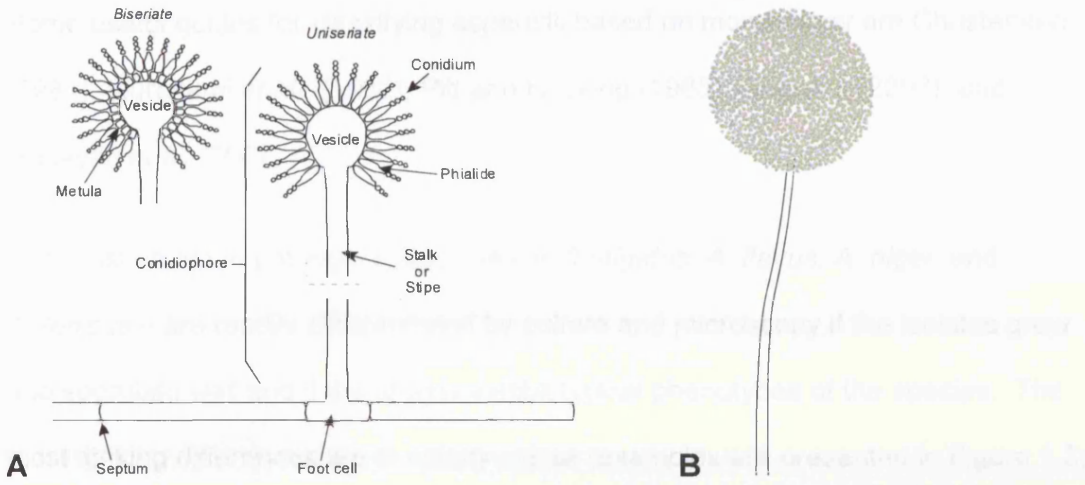
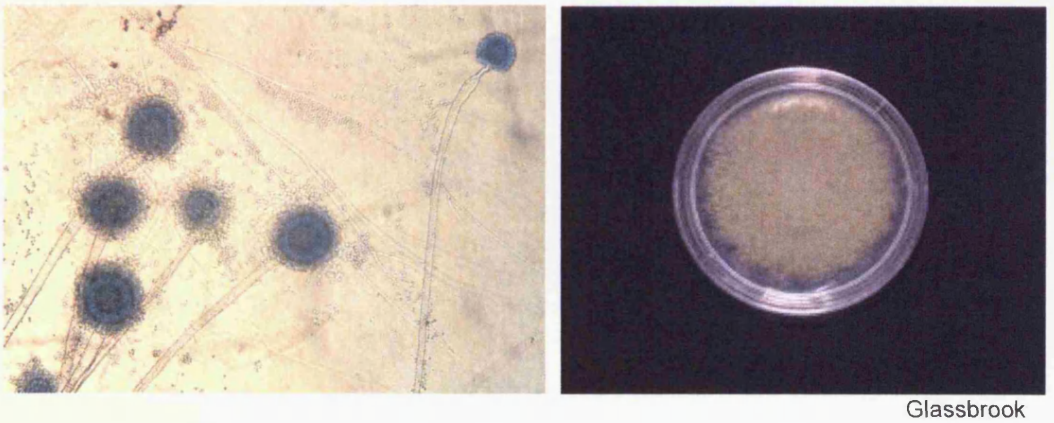


Figure 1.1. Typical conidial structures of *Aspergillus*. (A) Labelled cross-section of conidiophores and hyphae (redrawn/adapted from Deacon 1984 and Alexopoulos and Mims 1985). Note that the phialides, on which conidia form, may rest directly on the vesicle (uniseriate) or on another layer of cells that rest on the vesicle (biseriate). (B) general appearance of a typical *A. flavus* conidiophore under low magnification (Glassbrook).



Glassbrook

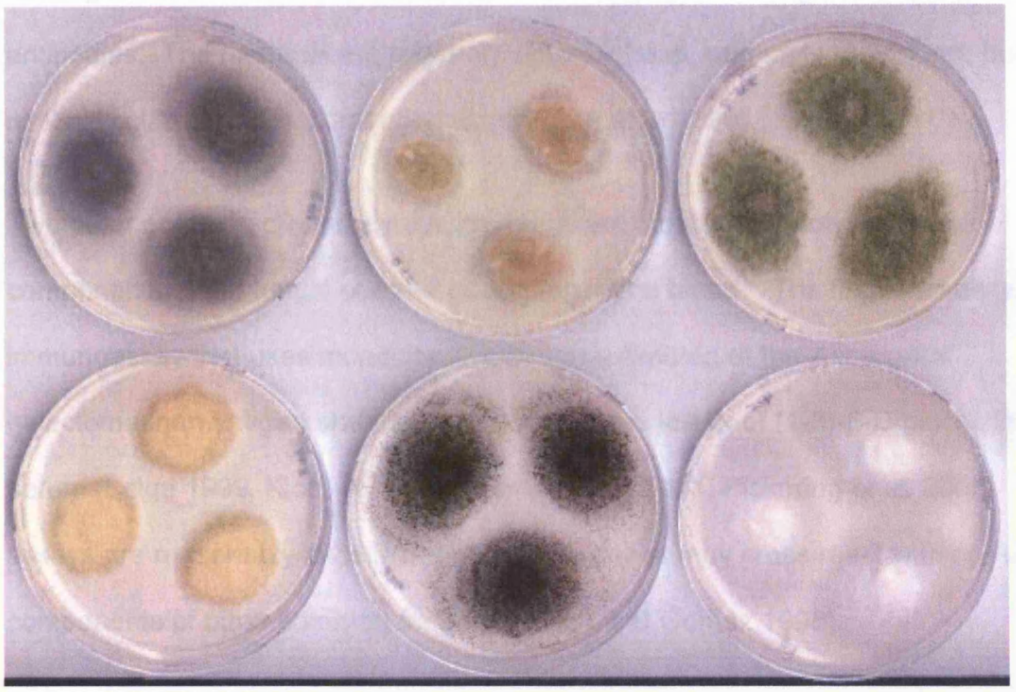
Figure 1.2. Condiophores and typical growth habit of *Aspergillus flavus*.

Some useful guides for classifying aspergilli based on morphology are Christensen (1981), Murakami *et al.* (1982), Pitt and Hocking (1985), Gugnani (2003), and Hedayati *et al.* (2007).

The most common pathogenic species - *A. fumigatus*, *A. flavus*, *A. niger*, and *A. terreus* – are readily differentiated by culture and microscopy if the isolates grow and sporulate well and if the strains exhibit typical phenotypes of the species. The most striking differences are in colony colour (examples are presented in Figure 1.3). *A. fumigatus* cultures are blue-grey, *A. flavus* yellow-green, *A. niger* black, and *A. terreus* cinnamon brown. However, identification by morphology and spore colour can be complicated by slow growth, poor sporulation or atypical characteristics of some field or clinical isolates (Denning 1998). Also, the culture procedure can take several days to over a week for the fungus to grow out, sporulate, and develop characteristic colour. In addition, closely related species may have very similar gross morphologies, sometimes making positive identification to species difficult. The general characteristics of the fungi included in this study are presented in Table 1.3.

Section *Circumdati* - biseriate, yellow to rust brown conidia, ochratoxins, ubiquinone CoQ-10(H_2)	
A. ochraceous	postharvest pathogen of agricultural products
Section *Flavi* - uniseriate and biseriate, pale yellow to olive green (typical) and brown conidia, +/- alflatoxins, kojic acid, ubiquinone CoQ-10(H_2)	
A. flavus	pathogenic; common contaminant of maize, peanuts, and tree nuts
A. oryzae	used in fermentations, closely related to *A. flavus*, morphology and colour varies widely
A. parasiticus	similar to *A. flavus*, not generally considered pathogen
A. sojae	used in fermentations
A. tamarii	used in fermentations
Section *Fumigati* – uniseriate, grey to blue green conidia, +/- gliotoxin, ubiquinone CoQ-10	
A. fumigatus	most prevalent *Aspergillus* pathogen of humans
Section *Nidulantes* – biseriate, blue green to deep green conidia, sterigmatocystin, ubiquinone CoQ-10(H_2)	
A. nidulans	sometimes pathogenic, widely used for genetic studies
Section *Nigri* – biseriate, dark brown to black conidia, large vesicles, +/-ochratoxin A, ubiquinone CoQ-9	
A. carbonarius	contaminant of coffee, grapes, and tree nuts
A. niger	rot pathogen of plants, sometimes mammalian pathogen, used in fermentations
Section *Terrei* - biseriate, tan to cinnamon brown, patulin, ubiquinones CoQ-10 and CoQ-10(H_2)	
A. terreus	emerging pathogen, capable of forming adventitious conidia (aleurioconidia)

Table 1.3. General classification and characteristics of *Aspergillus* species included in this study (Matsuda *et al.* 1992, Seifert and Levesque 2004, Varga *et al.* 2003). General morphological characteristics and some characteristic metabolites are listed for each section within the genus *Aspergillus*; individual species are listed within their respective sections.



Glassbrook

Figure 1.3. Examples of distinctive *Aspergillus* spore colours on a minimal medium. Top row, left to right: *A. fumigatus* NRRL163, *A. terreus* NRRL 255, *A. flavus* NRRL 3357. Bottom row, left to right: *A. ochraceous* NRRL 398, *A. niger* NRRL 326, *A. tamari* NRRL 425.

Immunological assays. Immunological techniques have been applied to detection and speciation of Aspergilli, but find more use in clinical settings as part of diagnosis for aspergillosis than as a routine tool for taxonomy. Assays are typically directed at components of the fungal cell wall or secreted proteins or, in immunocompetent individuals, at antibodies to fungal antigens (Latge 1999, Pasqualotto and Denning 2005).

Immunodiffusion and counterimmunoelectrophoresis are the most common serological methods used in the clinical laboratory for detecting anti-*Aspergillus* antibodies. The methods are relatively simple, cheap, and easy to perform, but are not particularly sensitive and not quantitative (Latge 1999).

Two other common clinical assays involve direct detection of polysaccharide components of the fungal cell wall circulating in the blood. The first is an enzyme immunoassay that uses monoclonal antibodies directed at the *Aspergillus* galactomannan antigen and the second measures levels of (1-3)-β-D-glucan in serum (Latge 1999, Klont *et al.* 2004, Hope *et al.* 2005, Pickering *et al.* 2005). These assays are reasonably sensitive, but in some cases may cross react with cell wall components of other common fungi (Hamilton and Gomez 1998).

Nucleic acid based techniques. As a group, Aspergilli are genetically well characterised compared to other fungi. Aspergilli have eight chromosomes and the genome sizes range from approximately 28 Mb to 38 Mb. A number of *Aspergillus* species have sequenced genomes, including *A. clavatus*, *A. flavus*, *A. oryzae*, *A. fumigatus*, *A. nidulans*, *A. niger*, and *A. terreus* (Galagan *et al.* 2005, Jones 2007, Kobayashi *et al.* 2007, Rokas *et al.* 2007, Hedayati *et al.* 2007). The availability of sequence data across *Aspergillus* species facilitates species-to-species comparisons at the DNA and protein levels.

Molecular techniques based on comparing DNA are emerging as a very powerful tool for classifying organisms which is much less subjective than the classical phenotypic approach. The analyses involve comparing the electrophoretic migration of entire chromosomes (karyotyping) or defined sections of DNA, hybridization, or directly comparing sequences (Clutterbuck 1994, McDonald 1997, Gil-Lamaignere *et al.* 2003). Examples of taxonomically differentiated genes often used for detection and identification, as well as determining phylogenetic relationships of Aspergilli and other

fungi are β-tubulin, calmodulin, cytochrome *b*, actin, hydrophobin, and ribosomal DNA (Frisvad *et al.*1998, Lutzoni et al. 2004, Samson *et al.* 2006, Yaguchi *et al.* 2007). These genes have been selected because as a rule they are universally present in these organisms and the contain stretches of highly conserved sequences that can be targeted with primers for amplification by polymerase chain reaction (PCR). The amplified sequences useful for taxonomy contain non-coding regions that tend to show more genetic variation than coding regions (Webster and Weber 2007). Examples of coding and non-coding regions in β-tubulin and ribosomal DNA are illustrated in Figure 1.4

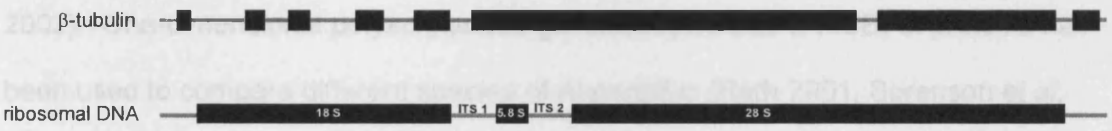


Figure 1.4. Examples of gene structures for β-tubulin and rDNA from *A. oryzae*. Coding regions are depicted as thick blank lines; non-coding regions are depicted as thinner grey lines. The internal transcribed spacers (ITS) flank the coding region for the 5.8 S RNA. Adapted from *A. oryzae* RIB40 sequences displayed in NCBI Sequence Viewer (http://www.ncbi.nlm.nih.gov/projects/sviewer/).

Ribosomal DNA (rDNA) is the most commonly used for taxonomic and phylogenetic studies because ribosomes are present in all cellular organisms, the sequences contain both variable and conserved regions, and primers are readily available for the different regions (Kurtzman 1994, Hillis and Dixon 1991). In addition, multiple copies of the ribosomal genes are present as tandem repeats in the genome and each haploid fungal genome has about 50-250 copies of the repeat that can be targeted for PCR (Webster and Weber 2007).

PCR provides a means to greatly amplify targeted sets of nucleic acid sequences. Nucleic acid methodologies based on PCR have the advantage of requiring very little starting material for analyses and thus can eliminate the time and sometimes difficulty required to culture the organism.

Protein profiling. Although not used extensively, protein profiles are an additional set of diagnostic features that can supplement other diagnostic approaches. The different protein patterns observed by electrophoretic techniques are directly related to the diversity of the coding genes and may indicate specific differences or similarities between species (Hennebert and Vancanneyt 1998, Mitterdorfer *et al.* 2002). One-dimensional polyacrylamide gel electrophoresis (PAGE) of proteins has been used to compare different species of *Aspergillus* (Rath 2001, Sorenson *et al.* 1971). Although different banding patterns show some similarities related to species, the technique, at this stage appears to be more useful for confirming identity or demonstrating similarities or differences rather than for directly identifying an organism. The electrophoretic patterns of isozymes (enzymes that catalyze the same reaction, but differ in amino acid sequence) have been used in some taxonomic studies with *Aspergillus* (Rinyu *et al.* 1995), and may provide additional features to support classification of an organism.

Chemical profiling. Aspergilli have a number of biochemical characteristics that, from a complete blind chemical profile, would immediately classify them as eukaryotic and as Eumycota or true fungi – they have cell walls containing chitin and glucan; they have sterols in their membranes, primarily ergosterol; the C16 and C18 chain lengths dominate the fatty acid profile; they produce trehalose and polyols (Wessels 2005).

Analysis of other chemical markers or patterns of metabolites are required for further classification.

Secondary metabolite profiles have been used to recognize individual species in conjunction with other approaches such as morphology and physiology (Guarro *et al.* 1999). The profiles of mycotoxins, particularly aflatoxins, are very important for classification of *Aspergillus* species (Klich *et al.* 2000, Frisvad *et al.* 1998, Seifert and Levesque 2004, Varga *et al.* 2004). A good overview of mycotoxins is presented by Bennett and Klich (2003).

Fatty acid profiles have been used extensively for chemotaxonomy of bacteria and the characterization of microbial communities (Kirk *et al.* 2004, Zelles 1999). Most methods involve preparing fatty acid methyl esters (FAME) and analyzing the FAME profile by GC or GC/MS. Fatty acid and lipid profiles have shown some promise in yeast taxonomy and identification (El Menyawi *et al.* 2000, Botha and Kock 1993). Although not extensively used, these techniques have been applied to identification of *Aspergillus* and *Penicillium* species (Nemec 1997, da Silva 1998).

Ubiquinone (coenzyme Q) is a lipid component of the mitochondrial electron transport chain and has been used to help classify a number of yeast and filamentous fungi (Paterson in Frisvad *et al.* 1998, Ahearn 1978). The number of isoprene units attached to the benzoquinone and the percentage of the reduced form present can support taxonomic grouping. Matsuda *et al.* (1992) report that three major ubiquinone systems (Q-9, Q-10, and Q-10(H2) were present in the *Aspergillus* species they studied and the ubiquinones were useful indicators for classification.

As with morphological techniques, most chemical profiling techniques require isolation and culturing of the organism under standardized conditions to be valid and may not be suitable for time sensitive identifications. They also typically require more

sample mass than molecular techniques based on PCR that can generate usable sequence from very tiny amounts of material. However, the common molecular markers used for identifying Aspergilli, such as β-tubulin and rDNA, may not adequately resolve the identification of an organism to the species level. Consequently, morphological and biochemical characterization is still often required to correctly identify an organism.

Mass Spectrometry Approach

In contrast to the targeted analyses that monitor specific compounds or groups of compounds as described above, biochemical profiling techniques based on mass spectrometry employ a few core methods to survey a wide variety of compounds simultaneously and the methods can be scaled to accommodate relatively small quantities of test material (Fiehn 2002, Glassbrook and Ryals 2001, Halket *et al.* 2005). This analytical approach was applied in this project to survey small molecule and protein compositions of different species of Aspergilli using gas or liquid chromatography coupled with mass spectrometry (GC/MS and LC/MS, respectively).

In GC/MS and LC/MS, chromatography is used to separate the components of sample extracts prior to detection by mass spectrometry. Chromatographic separation in the GC takes place as volatile components of the sample differentially partition between a carrier gas flowing through the chromatography column and a stationary liquid or polymer lining the column. Separations in the LC occur as the components of the sample partition between a liquid mobile phase and a stationary solid phase packing material in the column. In the mass spectrometer, molecules are ionized and then separated in electromagnetic fields based on their mass-to-charge ratio (*m/z*). The chromatographic retention, mass of a compound and the distinctive

masses of fragments from the compound (the spectrum) can be used to derive its identity. Profiling by GC/MS and LC/MS was recently reviewed by Halket *et al.* (2005).

GC/MS is well suited to the analysis of a wide range of biologically important compounds. However, because many of those compounds are not sufficiently volatile or chromatograph poorly by GC in their native states, they must be undergo chemical derivatization prior to analysis. In this study, crude extracts were treated to produce trimethylsilyl (TMS) derivatives of compounds with amine, hydroxyl, acid, or phosphate functional groups. This yields a very complex solvent extract suitable for GC/MS that contains a broad cross section of biologically important compounds, including TMS-derivatized alcohols, sugars, small organic acids, fatty acids, amino acids, and sterols. Many of these compounds are readily identified by matching electron impact (EI) spectra of the chromatographic peaks with spectra in the mass spectral library available from the National Institute of Standards and Technology (NIST). In many cases, the NIST library also lists calculated and/or experimentally measured retention indices to aid compound identification.

LC/MS is better suited for the analysis of heat labile compounds, high-molecular weight compounds, and non-volatile compounds in their natural form. In this study, LC/MS with electrospray ionisation (ESI) was used for analysis of small molecules and peptides from culture media and fungal homogenates. The term LC/MS is used as a generic term to describe analyse conducted with an HPLC coupled to a mass spectrometer. It may also be used, in context, to specifically describe analysis of ions as they are generated in the ion source and passed to the detector without fragmentation. The term LC/MS/MS is used to describe the process of isolating selected ions and fragmenting them to generate characteristic spectra. These spectra are used for selectively monitoring compounds or for further characterization

of a parent ion based on the fragmentation pattern, in a manner similar to the spectral matching used for the EI spectra generated in GC/MS. As with GC/MS the NIST spectral library includes some entries relevant to LC/MS/MS and may be used to prepare custom user libraries for storing MS/MS spectra of selected fungal metabolites such as mycotoxins. LC/MS/MS spectra of peptides may be matched to theoretical spectra generated from amino acid sequences in order to identify proteins contained in test samples.

Aspergillus, as a test organism, is particularly well suited for a biochemical profiling approach using GC/MS and LC/MS. Culture conditions and molecular biology techniques are well established. Some of the biochemical pathways have been extensively studied. Genomes of a good section of *Aspergillus* species have been sequenced and are publicly available. Differences between organisms observed in metabolic profiles, give clues to underlying differences in metabolism that may be traced back to changes in specific proteins, expression profiles, and specific genes. In addition, the availability of genomic data allows LC/MS identification of proteins based on the specific amino acid sequences encoded in an organism's genes. As with nucleic acid based techniques, there may be detectable amino acid sequences that are specific enough to a given species to allow identification based on those sequences. The ability to identify organisms based on sequence data will improve as the costs of genetic sequencing continues to decrease and more sequence data for more organisms become publicly available.

Objectives

The overall goal of this study was to characterise a cross section of *Aspergillus* species using LC/MS and GC/MS biochemical profiling techniques in order to find specific small molecules, peptides or biochemical profiles that differentiate the species. The initial focus was to catalogue small molecules and proteins from a number of reference strains with the expectation that clear patterns in chemical composition or unique peptide sequences would emerge that could be used to classify Aspergilli. Subsequently, techniques developed for characterizing the reference strains were applied in case studies to characterizing and identifying field isolates and to detecting the presence of fungi in host material. An additional objective of the case studies was to gather additional biochemical information on the organism in the context of the test systems studied by identifying the predominant enzymes and structural proteins present under different experimental conditions. In some cases, protein profiles of the fungus and the host material were obtained concurrently. The techniques developed and information gained from these biochemical profiles may serve as the basis for further studies on host pathogen interactions.

Chapter 2: Materials and Methods

General approach. Reference strains of different *Aspergillus* species, *Penicillium chrysogenum*, *Candida albicans* and *Cryptococcus neoformans* were grown under standardized growth conditions and then analyzed using GC/MS and LC/MS biochemical profiling techniques in order to find specific small molecules, peptides or biochemical profiles that differentiate the species. Subsequently, culture and morphology, biochemical profiling, and PCR were applied to characterize and identify several laboratory and field isolates. Protein profiling was also applied to detect the presence of fungi in host material (maize).

Chemicals and reagents. General laboratory chemicals were reagent grade or better and were purchased from various vendors, including Sigma-Aldrich (St. Louis, MO, USA), Fisher Scientific (Pittsburgh, PA, USA) and VWR International (West Chester, PA, USA). Basal Medium Eagle vitamins 100X stock solution and Basal Medium Eagle essential amino acids 50X stock solution were purchased from Sigma (St. Louis, MO, USA). Acetonitrile, isopropanol and methanol were chromatography grade, suitable for spectrophotometry, HPLC and GC analysis, and pesticide residue analysis (Honeywell Burdick & Jackson®, B&J Brand®). Deionized water (18 Mohm-cm) was purified with a NANOpure® DIamond™ ultrapure water system (Barnstead, Dubuque, IA, USA).

Fungal reference strains. Isolates of *Aspergillus* species, *Candida albicans*, and *Cryptococcus neoformans* used in this study are listed in Table 2.1. Most of the *Aspergillus* cultures were obtained from the United States Department of Agriculture (USDA) Northern Regional Research Laboratory (NRRL), now the National Center for Agricultural Utilization Research (NCAUR); Peoria, Illinois, USA), with the

Genus species	Strain	Notes
Aspergillus carbonarius (Bainier) Thom	NRRL 369*	Neotype strain of *A. carbonarius*
Aspergillus carbonarius (Bainier) Thom	NRRL 67	
Aspergillus carbonarius (Bainier) Thom	NRRL 346	
Aspergillus clavatus Desmazieres	NRRL 1*	Lectotype, sequenced strain
Aspergillus clavatus Desmazieres	NRRL 3538*	Military specification culture
Aspergillus flavus Link	NRRL 3357*	F.A. Hodges strain M52, SRRC 167, sequenced strain
Aspergillus flavus Link	NRRL 1957*	Neotype strain
Aspergillus flavus Link	NRRL 485*	isolated from insect
Aspergillus flavus Link	NRRL 3646	Military specification culture
Aspergillus flavus Link	Papa 827	white spores, *nor* mutant
Aspergillus fumigatus Fresenius	Af293*	Denning; sequenced strain
Aspergillus fumigatus Fresenius	NRRL 163*	Neotype strain *A. fumigatus*
Aspergillus nidulans(Eidam) G. Winter	FGSC A4*	Glasgow wild type
Aspergillus nidulans(Eidam) G. Winter	NRRL 187*	Neotype strain of *Emericella nidulans*, anamorphic state
Aspergillus niger van Tieghem	NRRL 326*	Neotype strain
Aspergillus niger van Tieghem	NRRL 3	
Aspergillus niger van Tieghem	NRRL 363	
Aspergillus ochraceous Wilhelm	NRRL 398	Ex-type strain
Aspergillus oryzae (Ahlburg) Cohn	NRRL 447	Neotype strain
Aspergillus oryzae var. *oryzae*	NRRL 1988	From soy sauce, syn. *A sojae*
Aspergillus flavus Link	NRRL 4823	Type strain *A. oryzae wehmeri*
Aspergillus oryzae (Ahlburg) Cohn var. *brunneus*	NRRL 5590*	syn. RIB 40; sequenced
Aspergillus oryzae (Ahlburg) Cohn var. *oryzae*	RIB128	National Research Institute of Brewing (Japan)
Aspergillus oryzae (Ahlburg) Cohn	RIB430	National Research Institute of Brewing (Japan)
Aspergillus parasiticus Speare	NRRL 424	Lectotype strain of *A. terricola* var. *americana*, naturally occurring brown mutant
Aspergillus parasiticus Speare	NRRL 502	Lectotype strain of *A. parasiticus*
Aspergillus parasiticus Speare	SU-1	Obtained from SRRC along w/sec- derived from SRRC 143A
Aspergillus tamarii Kita	NRRL 425	syn. *A. terricola*
Aspergillus terreus Thom	NRRL 255*	Lectotype strain
Aspergillus terreus Thom	NRRL 274*	Clinical isolate from human ear
Aspergillus terreus Thom	NRRL 1913*	Isolated from mouse lung
Aspergillus westerdijkiae Frisvad & Samson	NRRL3174	syn. *A. ochraceous* Wilhelm
Aspergillus westerdijkiae Frisvad & Samson	NRRL5175	syn. *A. ochraceous* Wilhelm
Cryptococcus neoformans (Sanfelice) Vuillemin (1901)	H99*	
Candida albicans (Robin) Berkhout (1923)	SC5314*	DUMC 115.01
Paecilomyces lilacinus	NRRL 895	
Penicillium chrysogenum	NRRL 807	Neotype strain of *P. chrysogenum*
Penicillium chrysogenum	NRRL 824	Fleming's penicillin producing strain
Petromyces alliaceous	NRRL 4181	Neotype strain of *P. alliaceus*
Petromyces alliaceous	NRRL 315	

Table 2.1. Fungal reference strains used in the study. All strains listed, except *A. tamarii* were analyzed by LC/MS/MS for protein; strains analyzed by GC/MS and LC/MS are marked with an asterisk. Note that project data for *A. tamarii* is referenced only in Figure 1.3 (spore colours).

following exceptions - *Aspergillus nidulans* A4 obtained from the Fungal Genetics Stock Center (FGSC; Kansas City, Missouri, USA), *Aspergillus fumigatus* Af293 obtained from Duke University Medical Center (Durham, North Carolina, USA), *Aspergillus flavus* Papa 827 obtained from Dr. Gary Payne (NC State University, Raleigh, North Carolina, USA), *Aspergillus parasiticus* SU-1 obtained from the USDA Southern Regional Research Center (SRRC; New Orleans, Louisiana, USA) and *Aspergillus oryzae* RIB128 and RIB430 were obtained from National Research Institute of Brewing (NRIB) in Japan. *Candida albicans* SC5314 and *Cryptococcus neoformans* H99 also were obtained from Duke University Medical Center (DUMC).

Fungal isolates for case studies. Field isolates were cultured from maize samples obtained at three different research stations in North Carolina (collected by Magen Starr Eller in James Holland's laboratory at NC State) and taken through an identification process based on morphological characteristics, mass spectrometry, and molecular techniques. Dried maize seeds, selected for discoloration characteristic of fungal infection, were obtained for the 2007 growth season at the Central Crops Research Station (Clayton, NC, USA), the Sandhills Research Station (Jackson Springs, NC, USA) and the Tidewater Research Station (Plymouth, NC, USA). Dry maize kernels were ground to a powder in a laboratory ball mill, then streaked onto agar-solidified BCP minimal medium (BCPmin, described below) as a powder and allowed to grow until mycelial growth was evident and *Aspergillus*-like reproductive structures were present. Colonies with the characteristic conidiophores bearing yellow to green conidia were subcultured to generate single conidia or hyphal tip isolates. Two non-*Aspergillus* fungi, one field isolate from maize and one lab contaminant from PDA were isolated to serve as an out group for comparison with *Aspergillus* strains. Isolates presumed to be *Aspergillus flavus* based on morphology were obtained and designated Clayton 270, Sandhills 174, and Plymouth 98 based on the source of the maize sample (research station and plot). Non-*Aspergillus*

isolates were designated as P98b (isolated from the same maize sample as the *Aspergillus* strain Plymouth 98) and GAPLAB_RH (lab contaminant from Rob Holmes in Gary Payne's laboratory at NC State). Summary information for the field isolates is presented in Table 2.2.

Designation	Tentative identification	Source
Clayton 270	*Aspergillus flavus*	maize kernels; Clayton NC, USA
Sandhills 174	*Aspergillus flavus*	maize kernels; Jackson Springs NC, USA
Plymouth 98	*Aspergillus flavus*	maize kernels; Clayton NC, USA
P98b	*Fusarium*	maize kernels; Clayton NC, USA
GAPLAB_RH	*Penicillium*	culture contaminant; Raleigh NC, USA

Table 2.2. Fungal isolates used in the case studies.

Growth Conditions. Fungi were grown in a chemically-defined minimal medium consisting of a balanced salt solution and trace vitamins with 1% w/v sucrose or 1% w/v glucose as the sole carbon source and 0.1%w/v ammonium chloride as the sole nitrogen source. The composition of the medium, designated BCPmin, is listed in Table 2.3. This composition of this defined medium was loosely based on various balanced salt solutions used for mammalian and insect tissue culture. Micronutrients were added, bicarbonate removed, and the phosphate concentration was adjusted to yield an unadjusted pH of 7.2+/-0.2 and to limit acidification of the medium by the fungus over three days of shake flask culture. The level of nitrogen (ammonium chloride) in the medium was set based on experiments with *Aspergillus flavus* and *Candida albicans* to optimize growth (fresh weight) at three days in shake flask culture at 28°C with various combinations of glucose and ammonium chloride. In these experiments, glucose ranged from 2 to 10 g/L and ammonium chloride ranged from 200 mg to 2 g/L in the balanced salt solution.

Ingredient	Formula	Amount		Concentration	
Sucrose	$C_{12}H_{22}O_{11}$	10	g/L	29.2	mM
Ammonium chloride	NH_4Cl	1	g/L	18.7	mM
Sodium chloride	NaCl	7.5	g/L	128	mM
Potassium chloride	KCl	300	mg/L	4.02	mM
Magnesium sulfate heptahydrate	$MgSO_4.7H_2O$	280	mg/L	1.14	mM
Calcium chloride dihydrate	$CaCl_2.2H_2O$	75	mg/L	510	µM
Sodium phosphate, dibasic, anhydrous	Na_2HPO_4	2.3	g/L	16.2	mM
Potassium phosphate, monobasic, anhydrous	KH_2PO_4	500	mg/L	3.67	mM
Ethylenediaminetetraacetic acid, disodium salt, dihydrate	$C_{10}H_{14}N_2O_8Na_2.2H_2O$	8	mg/L	21.5	µM
Ferrous sulfate, heptahydrate	$FeSO_4.7H_2O$	1.5	mg/L	5.40	µM
Zinc sulfate, heptahydrate	$ZnSO_4.7H_2O$	1.8	mg/L	6.26	µM
Copper sulfate, pentahydrate	$CuSO_4.5H_2O$	400	µg/L	1.60	µM
Manganese sulfate, monohydrate	$MnSO_4.H_2O$	200	µg/L	1.18	µM
Cobalt chloride, hexahydrate	$CoCl_2.6H_2O$	150	µg/L	0.63	µM
Boric acid	H_3BO_3	300	µg/L	4.85	µM
Sodium molybdate, dihydrate	$Na_2MoO_4.2H_2O$	300	µg/L	1.24	µM
Potassium iodide	KI	150	µg/L	0.90	µM
Agar for solidified medium		15	g/L		
Basal Medium Eagle vitamins 100X stock		1	mL/L		
Basal Medium Eagle essential amino acids 50X stock ,optional		2	mL/L		

Table 2.3. Composition of defined culture medium for culture of fungi for biochemical profiling studies (BCPmin).

The culture media were typically sterilized by autoclaving for 20 minutes at 121°C. Small volumes of media prepared for isotopic labelling experiments were sterilized by passing the medium through 0.22 µm nylon membrane filters (25 mm syringe or 50 mm vacuum type, sterile, Nalgene).

Shake flask cultures typically were incubated for three days in the dark at either 28°C or 37°C in a rotary shaker (100 rpm, New Brunswick G-25) using 500 mL baffled culture flasks containing 100 mL of medium at 1 x 10^6 conidia/mL. Cultures on agar-solidified BCPmin medium typically were incubated for 7-10 days on the bench top (approximately 22°C, ambient light) in 60 mm or 100 mm disposable culture plates or in 175 mL glass culture vessels (baby food jars) with Magenta® B-caps as closures.

Isolation of conidia and mycelia. Conidia harvested for fungal stocks were dislodged from cultures on solid medium with a stream of 0.05% v/v Triton X-100 in water by forcefully discharging the detergent solution against the culture surface. Conidia were left in this solution at 5°C for short term storage (up to a week). For long term storage (months to years), conidia were pelleted by centrifugation and re-suspended in 10% v/v glycerol in water for storage at 5°C or in 50% v/v glycerol in water for storage at -80°C. Conidia harvested for analysis were dislodged from cultures on solid medium with a stream of room temperature isopropanol; the conidial suspension was centrifuged to produce a pellet and the solvent decanted off; the conidial pellet was then gently air dried before processing.

Mycelia were harvested from liquid cultures by vacuum filtration through Miracloth (Calbiochem) backed with cotton gauze. Mycelia were pressed with laboratory wipers to remove residual liquid and then weighed in polypropylene scintillation vials and stored at -20°C or -80°C. Frozen mycelia were homogenized in a ball mill

(MiniBeadbeater-8™, BioSpec Products, Inc.) with ice cold 60% v/v methanol in water, typically for one minute with a mixture of beads ranging from 0.5 mm to 2.3 mm, to obtain 200-250 mg/mL of ground mycelia in homogenate. This homogenate was used as the starting material for small molecule analysis and protein extraction. Methanol/water homogenates were used as a common starting point for analyses because the homogenates remained fluid at -20°C and multiple aliquots could be removed for different types of analyses without introducing a freeze/thaw steps that might compromise sample integrity over time.

Test samples for case studies. Isolates were cultured on agar-solidified minimal medium in 60 mm disposable culture dishes for four days to produce biomass for analysis by mass spectrometry or for DNA isolation and sequencing of the ITS region of rDNA. Biomass was harvested by flooding the cultures with 5 mL of ethanol and scraping fungal material from the plate into the alcohol, and then the biomass was collected by briefly centrifuging the sample and discarding the supernatant. LC/MS/MS analyses for aflatoxin production were conducted on liquid minimal medium, methanol/water extracts of agar-solidified potato dextrose, or methanol/water extracts of infected maize kernels and compared to aflatoxin production by the reference strain *Aspergillus flavus* NRRL 3357. Solid samples were homogenized in cold 60% aqueous methanol (-20°C) with a laboratory ball mill (MiniBeadbeater-8™, BioSpec Products, Inc.) prior to extraction of protein, extraction DNA, or analysis of aflatoxin.

Apoplastic wash fluids (AWF) were isolated from uninfected control and *Aspergillus flavus*-infected corn kernels based on vacuum infiltration and centrifugation techniques described by Rohringer *et al.* (1983) and Lohaus *et al.* (2001). The procedures yielded approximately 50 μL of apoplastic washing fluid (AWF) from

4 maize kernels (samples kindly provided by Andrea Dolezal and Crystal Phelps in Gary Payne's laboratory).

GC/MS analysis. 100 μL subsamples of fungal homogenate, each equivalent to 20 mg fresh weight of biomass, were transferred into 2 mL instrument vials with 500 μL of isopropanol and a small volume of Teflon powder (a chemically inert support that aids with resuspending residue after drying). The homogenates were then reduced to dryness by vacuum centrifugation (Savant SpeedVac, SVC200H). The residue was resuspended in 100 μL of acetonitrile and 25 μL of a derivatizing reagent composed of 90% N,N-dimethyltrimethylsilylamine (TMS-DMA) and 10% hexamethyldisilazane (HMDS) then incubated at approximately 60°C for an hour. The solvent contained 50 mM trifluoroacetic acid and 100 mM N-methylmorpholine (aids derivatization of some organic acids and limits rearrangement of sugars) along with 9μg/mL dibutylpyridine, 6μg/mL 4-butylphenol and 3 μg/mL octyl-β-D-glucoside as internal standards (Glassbrook and Deighton, unpublished). The resulting extracts were centrifuged, transferred to small volume glass inserts in 2-mL glass instrument vials and analyzed by GC/MS without further cleanup.

GC/MS was performed with a Thermo Trace GC Plus gas chromatograph coupled to a Thermo DSQII mass spectrometer equipped with an electron impact ion source. Chromatographic separations were achieved with a Restek Rtx®5Sil MS (5% phenyl polysiloxane) column (30 m; 0.25 mm I.D.; 0.25 μm film thickness). Helium, the carrier gas, was set at a constant flow of 1 mL/min (linear velocity of ~40 cm/s). One microliter of extract was injected into a PTV injector operated in splitless mode at 160°C which was raised immediately at 5 °C/s to 360°C, held at 360°C for 3 min, lowered at 2 °C/s to 240°C, then held at 240°C. The column oven temperature was

programmed for a 60°C initial temperature with an 8°C/min ramp to a final temperature of 360°C (Glassbrook and Deighton, unpublished).

A retention standard containing 10 μg/mL each of even numbered n-alkanes ranging in length from C10 to C40 for calculating retention indices was prepared by diluting a Florida TRPH reference standard (Restek) in a solvent mixture of three parts isooctane, one part tetrahydrofuran, and one part benzene. A composite sample was prepared by combining equal aliquots of homogenates from all the samples in a sample set. The composite sample was used as a reference sample to verify linearity for sample analytes and to generate a list of target compounds for sample analysis. In a manner similar to external calibration with reference standards, aliquots of the composite equivalent to fresh weights of 5 mg to 30 mg biomass were analyzed concurrently with the test samples. Sample components that yielded reproducible detector responses with positive linear correlation ($r \geq 0.9$) with sample loading were considered valid for relative quantitation and sample to sample comparisons.

Chromatographic peaks were catalogued based on retention index (Kovats) and spectra. The retention indices were calculated by linear interpolation from chromatographic retention times relative to those of the reference alkanes. Relative quantitation was conducted by comparing peak areas in the fungal samples with to those of a composite reference sample at different sample loadings analyzed concurrently with the fungal extracts. The chromatographic data were processed using Xcalibur® 1.4 software (Thermo) and AnalyzerPro™ 1.1 from Spectral Works then exported to Microsoft® Office 2003 Excel® for sorting, basic calculations and graphing. An Excel® add-in, statistiXL Version 1.8, was used for statistical functions such as discriminant analysis.

LC/MS analysis. For small molecule analysis, aliquots of the homogenates were centrifuged and the supernatants submitted for LC/MS analysis; liquid samples (culture medium) were centrifuged to remove particulate and submitted for analysis. LC/MS analyses were performed with a Thermo Surveyor liquid chromatograph coupled to a Thermo LTQ linear ion trap mass spectrometer. Chromatographic separations were achieved with a C18 HPLC column (typically Thermo Hypersil Gold: 150mm x 2.1mm I.D., 5μm particle size, 175Å pore size). The mobile phase composition was programmed with a solvent gradient, initial conditions of 250 μL/min with a mixture of 90% acidified water (50mM acetic acid) and 10% acetonitrile programmed with a linear ramp to a final mixture of 10% acetonitrile and 90% isopropanol over the course of 18 minutes (Glassbrook and Deighton, unpublished). Although most LC/MS methodologies employ acetonitrile as the solvent, isopropanol was selected for this study in order to ensure that the water/solvent mixture was sufficiently non-polar to elute the bulk of lipids from the C18 column during the chromatographic run. Acetonitrile, present at a constant 10%, limits the pressure build up as the proportion of isopropanol increases throughout the gradient.

Derivatives of morpholine and hydroxyethylpyrrolidone were evaluated as retention standards for LC/MS. The retention standards were prepared by reacting morpholine and hydroxyethyl pyrrolidone with acid chlorides (acetyl chloride to palmitoyl chloride) in the presence of sodium carbonate or an organic base. A small volume of the reaction mixture was reduced to dryness and dissolved in mobile phase without further cleanup prior to analysis.

Chromatographic peaks were catalogued based on retention time, retention indices and spectra. Relative quantitation was conducted by comparing peak areas in the fungal samples with to those of a composite reference sample at different sample

loadings analyzed concurrently with the fungal extracts. The chromatographic data were processed using Xcalibur® 2.0 software (Thermo).

Aflatoxin analysis, Liquid samples (culture medium) or homogenates of solid samples (200 mg fresh weight per millilitre in 60% methanol by volume in water) were centrifuged to remove particulate and submitted for analysis. Chromatographic separations were achieved with a C18 HPLC column (typically Thermo Hypersil Gold: 150mm x 2.1mm I.D., 5um particle size, 175A pore size) running isocratically at 250uL/min with a mixture of 75% acidified water (50mM acetic acid),10% acetonitrile, and 15%isopropanol. The mass spectrometer was operated in positive mode with an electrospray ionization source and collecting MS/MS spectra for *m/z* 313 $[M+H]^{+1}$ (aflatoxin B1) and *m/z* 315 $[M+H]^{+1}$ (aflatoxin B2).

Quantitation of aflatoxin was conducted by comparing peak areas obtained for the MS/MS product ions of aflatoxin B1 and B2 (*m/z* 285 and 287 respectively) with those of a series of reference standards analyzed concurrently with the samples. The chromatographic data were processed using Xcalibur® 2.0 software (Thermo).

Protein extraction. Extracts of reference strains were prepared based on a diethylamine extraction described by Nolan and Teller (2006) and adapted for direct extraction of proteins from the methanol/water homogenates (Glassbrook, unpublished). N-methylpyrrolidone and N-octylpyrrolidone were added to enhance extraction of the protein and aid re-dissolution of the dried residue. Hydroxyethyldisulfide was used as a cysteine blocking reagent. All components were chosen for compatibility with downstream Bradford protein assay, LC/MS and SDS-PAGE.

For protein extractions of reference strains, 200 µL of mycelial or spore homogenates were transferred to a 1.5 mL microcentrifuge tube and precipitated with 500 µL of cold isopropanol. After centrifugation for 2 minutes at 10,000 x g (Eppendorf Microcentrifuge, Model 5415D), the supernatant was removed and discarded. The pellet was then resuspended in 100 µL of 60:40 methanol:water and precipitated again with 200 µL of isopropanol. After centrifugation for 2 minutes at 10,000 x g, the supernatant was removed and discarded. The pellet was allowed to air dry briefly before resuspending the pellet in 100 µL of a protein extraction solution (2% v/v diethylamine, 5% v/v N-methylpyrrolidone, 0.1% v/v N-octylpyrrolidone, and 50 mM hydroxyethyldisulfide; referred to as CxP/DEA/HED) and heating the sample for at least 15 minutes at 60°C. After incubation, the sample was centrifuged for 2 minutes at 10,000 x g and a 10 µL aliquot of the supernatant was mixed with 500 µL of Bradford reagent to assay for protein content. An aliquot of the protein extract containing 50-100ug of protein was reduced to near dryness by vacuum centrifugation. The protein residue was redissolved in loading buffer for SDS-PAGE or digested with trypsin prior to LC/MS analysis. Trypsin digests were prepared by resuspending 50-100 µg of protein in 100 µL of buffer containing 100 mM N-methylmorpholine and 50 mM acetic acid (pH 7.4) and adding 1 µg of trypsin (proteomics sequencing grade, Sigma) in 25 µL of 50 mM acetic acid. The proteins were digested for at least two hours at 37°C before analysis by LC/MS/MS.

For protein analysis of case study samples, the homogenates from fungal cultures that remained after DNA extraction, described below, were processed in a manner similar to that described for the reference strains, except that proteins were extracted directly from homogenates on methanol/water without a precipitation step. The remaining homogenates (~400 µL) were mixed with 45 µL of a 10X stock of CxP/DEA/HED (50% v/v N-methylpyrrolidone, 1% v/v N-octylpyrrolidone, 20% v/v diethylamine, and 500 mM hydroxyethyl disulfide in water). The samples were then

incubated at 60°C for one hour, reduced to dryness before trypsin digestion by redissolving the residue in 100 µL NMMA buffer, pH 7.4, along with 1 µg of trypsin in 25 µL of 50 mM acetic acid and then incubating the samples overnight at 37°C. LC/MS/MS analysis was as previously described for analysis of the reference strains.

The AWF protein samples were prepared for protein analysis by adding approximately 6 µL of a 10X stock of CxP/DEA/HED. The samples were then incubated at 60°C for one hour, reduced to dryness before trypsin digestion by redissolving the residue in 75 µL NMMA buffer, pH 7.4, along with 1 µg of trypsin in 25 µL of 50 mM acetic acid and then incubating the samples overnight at 37°C. LC/MS/MS analysis was as previously described for analysis of the reference strains.

LC/MS/MS analyses. Peptide analyses, were performed with a Thermo Surveyor liquid chromatograph coupled to a Thermo LTQ linear ion trap mass spectrometer. Chromatographic separations were achieved with a C18 HPLC column (typically Thermo Hypersil Gold: 150 mm x 1 mm I.D., 5 µm particle size, 175Å pore size). The mobile phase composition was programmed with a solvent gradient, initial conditions of 50 µL/min with a mixture of 95% acidified water (50 mM acetic acid) and 5% acetonitrile programmed with a linear ramp to a final mixture of 40% acidified water and 60% acetonitrile over the course of 55 minutes. The mass spectrometer was operated in data dependent MS/MS scan mode scanning from *m/z* 420-2000 and collecting MS/MS spectra on the four most abundant ions in each scan. Instrumental conditions for peptide analysis were developed in the Metabolomics and Proteomics Laboratory (now the Genomic Sciences Laboratory) at NC State University (Glassbrook and Deighton, unpublished).

Identification of proteins in the extract was performed by searching tandem mass spectra against protein databases in FASTA format using Bioworks™ Browser

software (version 3.3, Thermo). Result files from BioWorks™ were exported to Scaffold™ (version 2.02.01, Proteome Software Inc., Portland, OR) and tandem mass spectra were processed using an additional algorithm, X!Tandem (www.thegpm.org; version 2007.01.01.1). Cysteine mercaptoethanol of cysteine (+76) was specified in Sequest and X!Tandem as a fixed modification.

Scaffold™ was used to validate MS/MS based peptide and protein identifications in the compiled results. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (Keller *et al.* 2002). Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii 2003). FASTA databases were compiled locally from subsets of protein databases downloaded from the National Center for Biotechnology Information (NCBI) or UniProt Knowledgebase (UniProtKB, Universal Protein Resource). Amino acid sequence alignments were prepared with ClustalX, version2.0.10 (described in Larkin *et al.* 2007) or with various online versions of Clustal. Clustal dendrograms were viewed and formatted with TreeView, version 1.6.6.

DNA extraction, PCR and sequencing. ITS sequencing was conducted on genomic DNA extracted with a CTAB methodology adapted from He *et al.* (2007). In brief, the fungal material from 4 day old cultures was homogenized with 25 mg polyvinylpolypyrrolidone (PVPP) in 500 µL of cold 60% v/v methanol/water (-20°C), and then a 100 µL aliquot of the homogenate was mixed with 400 µL of CTAB buffer and 10 µL of β-mercaptoethanol, incubated for 15 minutes at 60°C, partitioned against 500 µL chloroform and a 250 µL aliquot of the upper aqueous layer transferred to a clean microcentrifuge tube for precipitation with ammonium acetate and isopropanol. The resulting DNA pellet was rinsed with two, 200 µL volumes of

cold 70% ethanol and dried briefly (less than ten minutes) by vacuum centrifugation before redissolving the DNA in 50 µL of Tris/EDTA buffer, pH 8.

A 5 µL aliquot of the DNA solution was used as a template for PCR amplification using fungal ITS region primers (ITS1: 5'-TCC GTA GGT GAA CCT GCG G-3'; ITS4: 5'-TCC TCC GCT TAT TGA TAT GC-3') described by White *et al.* (1990) supplied by Integrated DNA Technologies (Coralville, IA, USA) in a 50 µL reaction volume (reagents from Takara Bio USA). Thermocycler conditions were as follows: initial denaturation 95°C for 5 minutes followed by 38 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 1 minute, followed by a final extension at 72°C for 5 minutes and a hold temperature of 4°C after the final extension. PCR products were separated on 0.8% agarose gels and visualized with ethidium bromide by UV illumination to verify the presence and size of the products prior to sequencing. PCR products were purified and taken to a 50 µL final volume using QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA, USA) in accordance with the manufacturer's instructions. DNA was sequenced at a commercial facility (GENEWIZ, Inc., South Plainfield, NJ, USA) using ITS1 and ITS4 primers. Multiple sequence alignments of ITS regions were performed with Clustal version 2.

Chapter 3: Results of GC/MS Analyses

Gas chromatography coupled with mass spectrometry (GC/MS) was used in this study to conduct small molecule profiling of fungal samples to look for compounds unique to different species of Aspergilli and to evaluate clustering as a tool for speciation. Chemical components were catalogued by chromatographic retention (retention time and retention index) and mass spectra. Profiling of biological samples by GC/MS yields a wide cross section of small molecules including TMS-derivatized alcohols, sugars, small organic acids, fatty acids, amino acids, and sterols. Typical profiling analyses of biological samples produce semi-quantitative data for hundreds of compounds. However, fungal cultures produced on common defined laboratory media, particularly media used specifically for aflatoxin production, are atypical in that a few polyols dominate the resulting profiles and the number of components detected in each sample is on the order of one hundred, not several hundred. In addition to GC/MS profiling, some targeted analyses (directed at specific compounds or groups of compounds) were attempted in this study to collect data on a different cross section of components, not dominated by polyols. Analyses targeted on lipids yielded data on hydrophobic compounds such as fats, fatty acids, and sterols.

Profiling of TMS-Derviatized Samples. In this study, the analytical conditions produced chromatographic peaks with Kovat retention indices in the range of 1000 (C10) to greater than 4000 (C40). Of approximately 100 peaks detected, 23 chromatographic peaks, representing 21 compounds, passed basic reproducibility and linearity checks (Table 3.1). An example of sample matrix loading (composite equivalent fresh weight) versus detector response (area ratio of analyte to internal standard) for erythritol is presented in Figure 3.1.

Compound	Retention time	Retention index	*m/z*	Abundance
Valine	9.60	1226	144, 218	0.13
Glycerol	10.50	1279	205, 218	11.8
iloleucine	10.96	1306	158, 213	0.07
Proline	11.14	1316	142	0.09
Succinic acid	11.45	1335	147, 247	0.15
Serine	12.04	1369	204, 218	0.30
Threonine	12.43	1392	117, 218, 291	0.22
Malic acid	14.14	1501	233, 245	0.12
Erythritol	14.26	1509	103, 205, 217, 307	10.2
Methionine	14.78	1543	128, 176	0.02
Pyroglutamic acid	14.92	1552	156	0.22
Arabitol	17.30	1716	103, 205, 307, 319	43.7
Fructose1	18.44	1797	217, 437	2.29
Fructose2	18.55	1806	217, 437	2.23
Glucose1	19.65	1892	191, 204, 217	3.30
Mannitol	20.04	1923	103, 205, 319	100
Glucose2	20.73	1978	191, 204, 217	5.79
Inositol	22.00	2085	305, 318	0.05
Linoleic acid	23.85	2249	337	0.11
Stearic acid	24.20	2282	117, 341	0.18
Trehalose	28.56	2724	191, 361	2.06
Squalene	29.90	2876	69, 81	0.04
Ergosterol	33.45	3322	337.5, 363.5, 468.5	0.33

Table 3.1. A subset of analytes evaluated for use in classification of fungi. The retention indices are for GC/MS analyses conducted with a 5% phenyl polysiloxane phase and programmed oven temperature. Abundances are reported for the composite sample and are normalized to total area counts for mannitol.

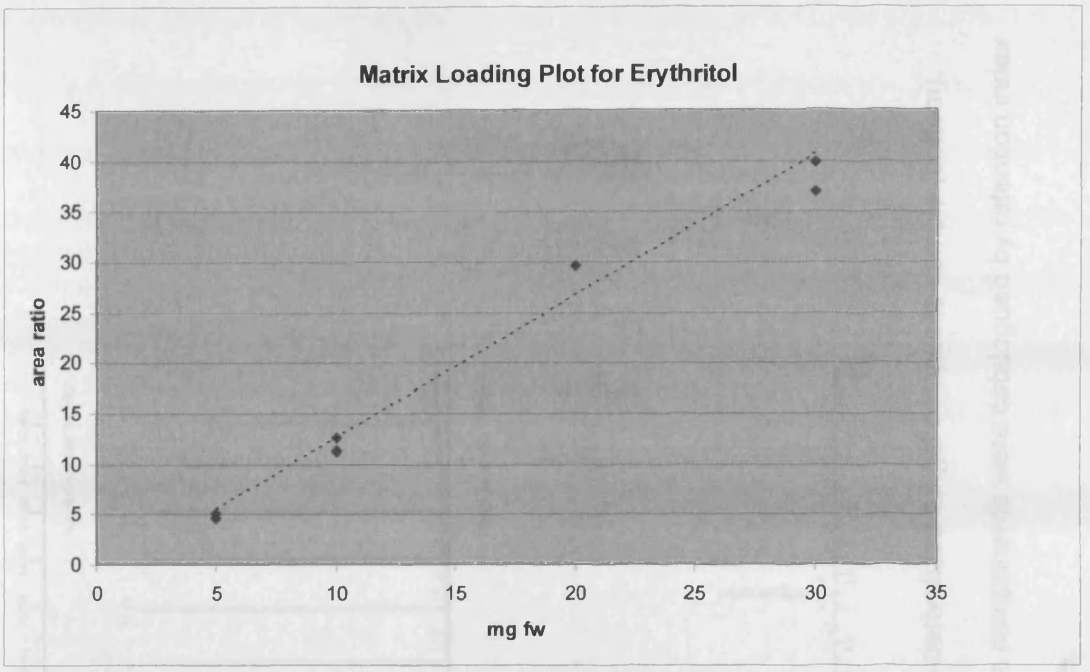


Figure 3.1. Example of a matrix loading plot for erythritol in fungal tissue analyzed by GC/MS profiling. The linear regression line is presented as a dashed line. Area ratio is the area counts for the analyte divided by the area counts for the internal standard, octyl-β-D-glucoside. Note that the range for matrix loadings evaluated for the composite sample brackets the 20 mg fresh weight (fw) loading of the test samples.

As expected in fungal samples (mycelia and cells), the major fatty acids were palmitic acid (C16:0) and a mixture of stearic acid (C18:0) and unsaturated fatty acids; ergosterol was the major sterol detected. Glycerol, erythritol, arabitol, and mannitol were the major polyols; trehalose was the major storage carbohydrate; succinate and malate were the major small organic acids; valine, isoleucine, proline, serine, threonine, methionine and glutamic acid (as pyroglutamic acid) were the major detected amino acids. Typical chromatograms of the paraffin mix retention standard and a TMS-derivatized composite sample are presented in Figure 3.2.

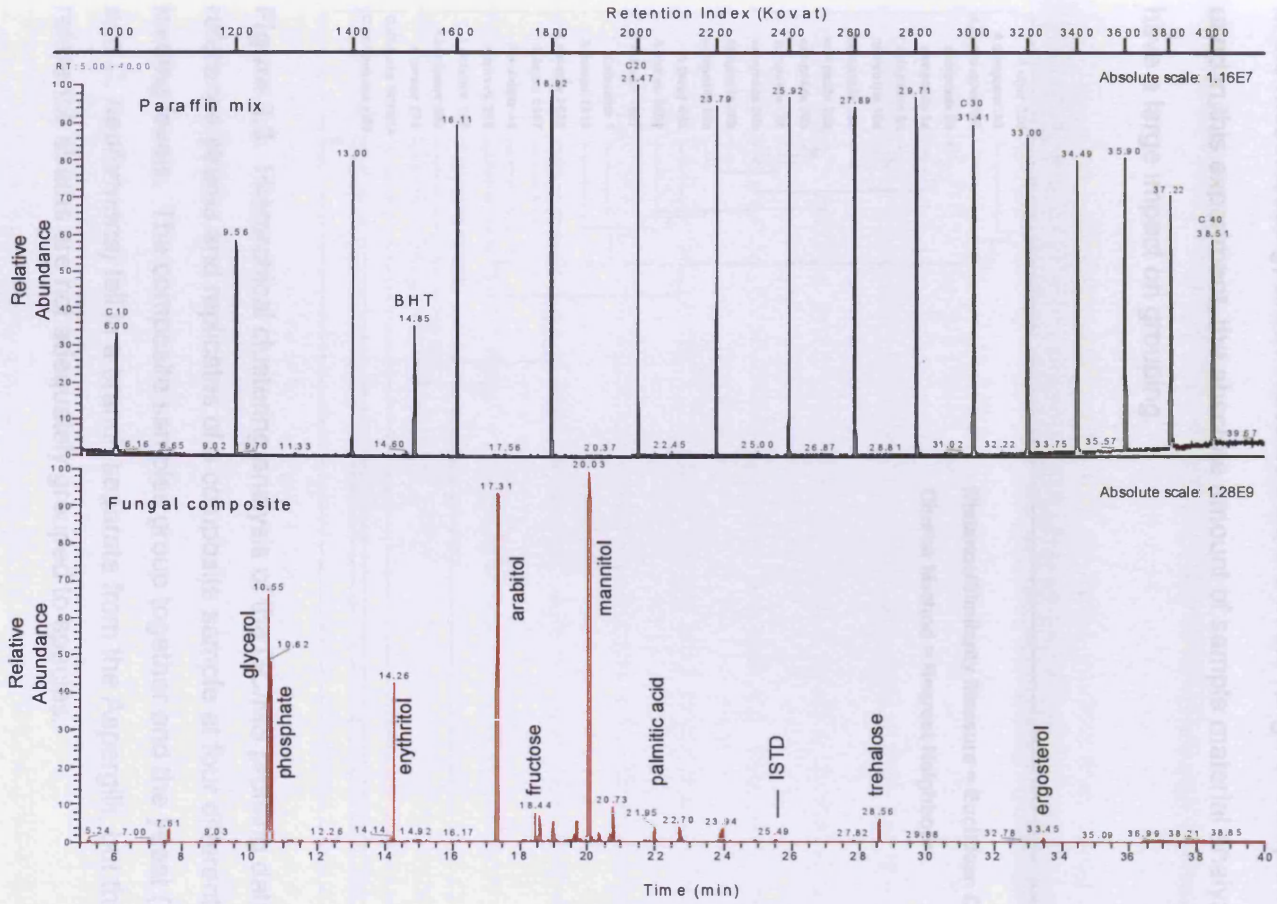


Figure 3.2. Typical chromatograms of a paraffin mix retention standard (top) and a TMS-derivatized fungal composite sample (bottom). Note the retention time and retention index scales on the horizontal axes. Chromatographic components were catalogued by retention index and spectra and maintained in a user library under the NIST Mass Spectral Search Program.

Hierarchical clustering analysis was applied to the subset of analytes presented in Figure 3.1 as an exploratory technique to determine if general patterns in the chemical profiles would group unclassified samples into taxonomically meaningful clusters. This unsupervised classification approach yielded mixed results as shown in Figure 3.3. The composite samples grouped together, as they should, regardless of the matrix loading, which indicates that, within the range of sample matrix loadings used in this experiment, the absolute amount of sample material analyzed does not have a large impact on grouping.

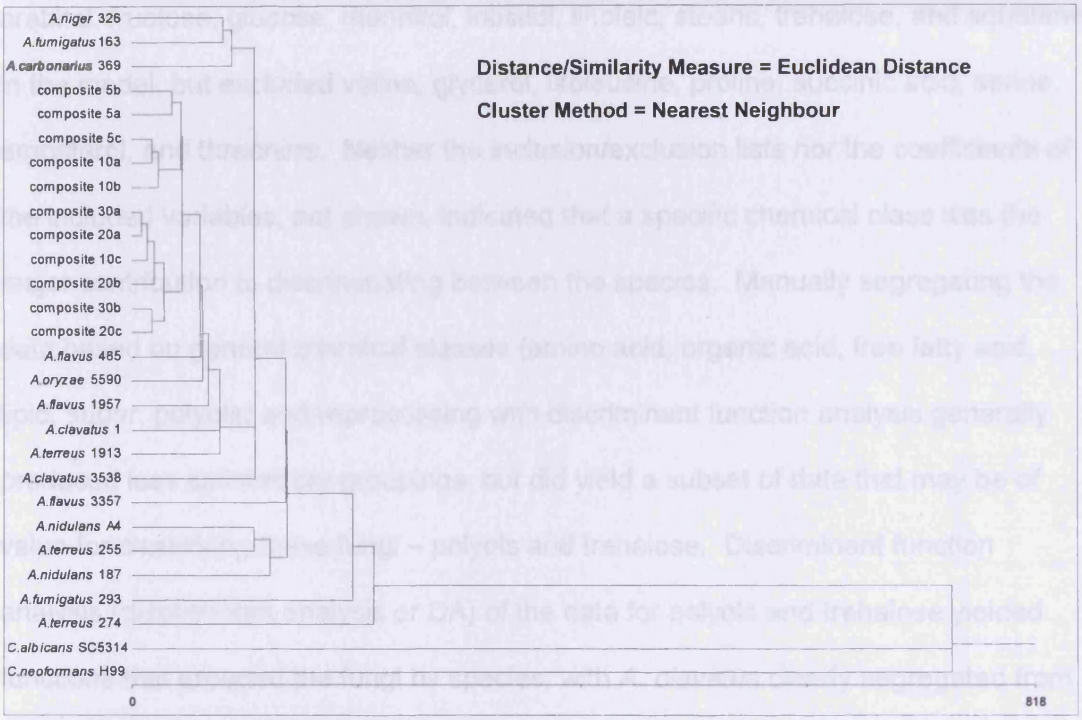


Figure 3.3. Hierarchical clustering analysis of the GC/MS profiling data set for 17 reference strains and replicates of a composite sample at four different matrix loading levels. The composite samples group together and the yeast (*C. albicans and C. neoformans*) fall in a branch separate from the Aspergilli, but the *Aspergillus* reference strains are not adequately grouped to species.

Discriminant function analysis, with samples labelled as species or as composite, was performed on the profiling subset with tolerances set to reject variables with tolerance <0.001. This yielded classifications with replicates of species in relatively tight groups with good separation from other species (Figure 3.4). Species from Section *Flavi* (*A. oryzae* and *A. flavus*) and Section *Nigri* (*A. carbonarius* and *A. niger*) are closely related genetically and did group together, but *A. fumigatus* and *A. clavatus* are also closely related and were well separated.

The tolerance setting allowed malic acid, erythritol, methionine, pyroglutamic, arabitol, fructose, glucose, mannitol, inositol, linoleic, stearic, trehalose, and squalene in the model, but excluded valine, glycerol, isoleucine, proline, succinic acid, serine, ergosterol, and threonine. Neither the inclusion/exclusion lists nor the coefficients of the included variables, not shown, indicated that a specific chemical class was the major contribution to discriminating between the species. Manually segregating the data based on general chemical classes (amino acid, organic acid, free fatty acid, lipid, sugar, polyols) and reprocessing with discriminant function analysis generally produced less satisfactory groupings, but did yield a subset of data that may be of value for classifying these fungi – polyols and trehalose. Discriminant function analysis (discriminant analysis or DA) of the data for polyols and trehalose yielded functions that grouped the fungi by species, with *A. clavatus* clearly segregated from the other Aspergilli (Figure 3.5). Glycerol was not a good predictor for classifying the fungi. Data for erythritol, arabitol, mannitol and trehalose had significantly different means across the different species (groups), so these data were used as the weighted variables to derive functions that allowed discrimination between the different species.

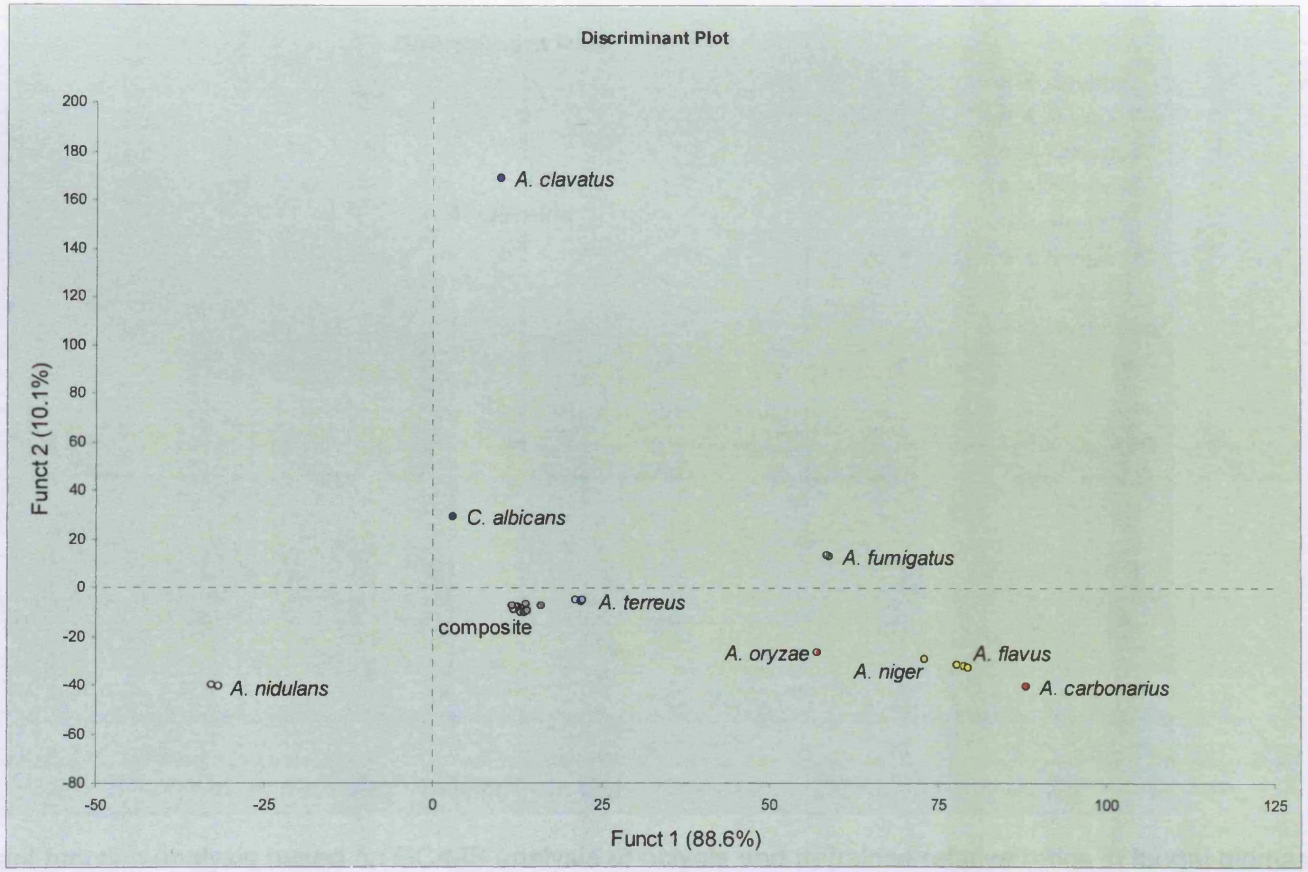


Figure 3.4. Discrimant function analysis of the GC/MS profiling data set (21 compounds) for 17 reference strains and composite samples. Supervised classification yielded clear species groupings of reference strains. *A. flavus*, *A. oryzae*, *A. niger* and *A. carbonarius* are closely related genetically and group together, but *A. clavatus* and *A. fumigatus*, also closely related, do not. Note that *C. neoformans* is off scale left (-744, -11.7).

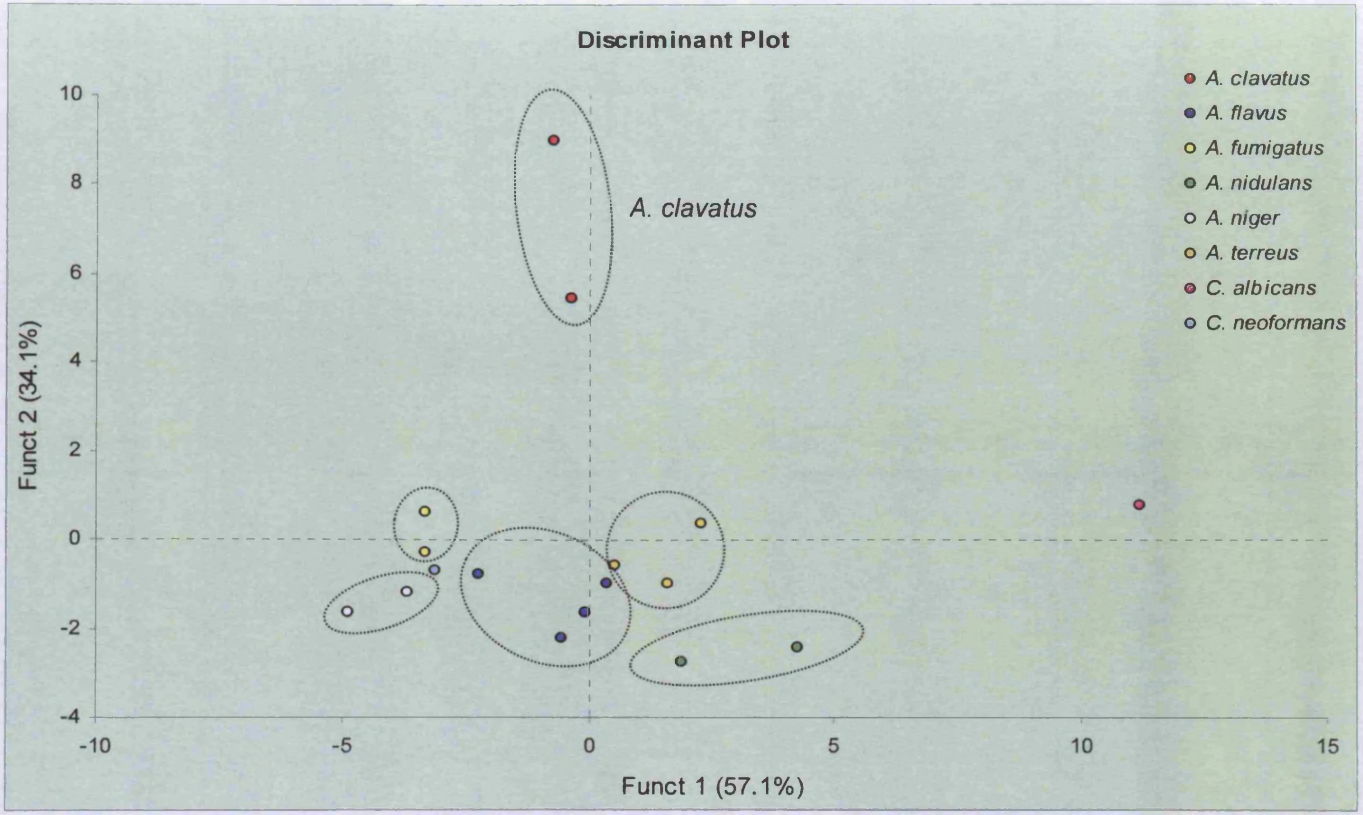


Figure 3.5. Discriminant function analysis based on GC/MS analysis of polyols and trehalose relative ratios in fungal biomass. Fungi can be grouped appropriately based on polyol-trehaolse profiles and *A. clavatus* (labelled) clearly segregates from the other fungi.

In a follow up analysis of the polyol-trehalose profiles, the relative proportions of the dominant polyols along with the disaccharide trehalose were somewhat diagnostic for classification (Figure 3.6).

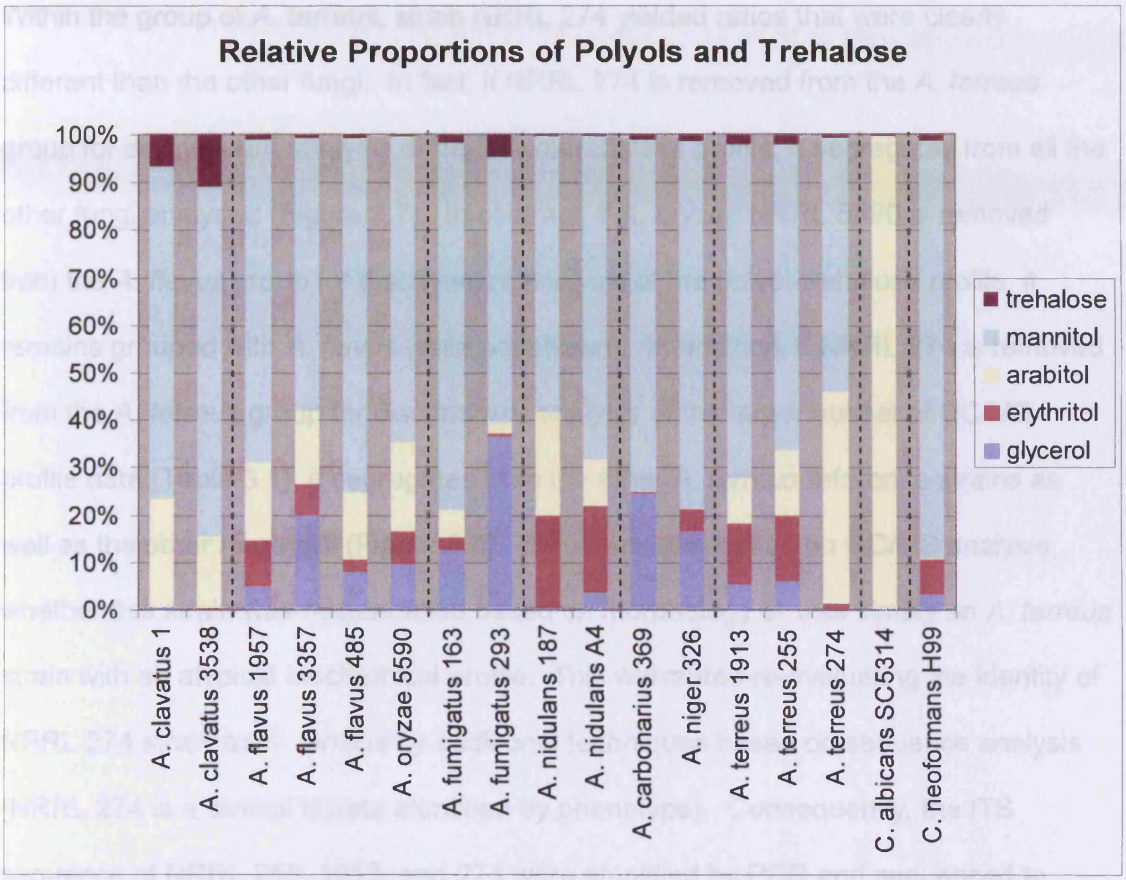


Figure 3.6. Normalized levels of polyol and trehalose in fungal samples expressed as the percentage each compound contributes to the total abundance (area ratio of analyte to internal standard) in each sample, one replicate per strain. Data is sorted based on species.

Mannitol was generally the most abundant polyol detected in the Aspergilli and in *Cryptococcus neoformans*. Arabitol was the dominant polyol in *Candida albicans*. There were some patterns evident in the relative ratios of polyols and trehalose that tended to group species. For example, *A. clavatus* tended to yield mannitol at the highest relative percentage followed by arabitol and trehalose, but little glycerol or erythritol. *A. carbonarius* and *A. niger* yielded mannitol at the highest relative

percentage followed by glycerol, with little very little arabitol. The other Aspergilli produced mannitol at the highest relative percentage along with various levels of the other polyols and trehalose.

Within the group of *A. terreus*, strain NRRL 274 yielded ratios that were clearly different than the other fungi. In fact, if NRRL 274 is removed from the *A. terreus* group for discriminant analysis of the polyol-trehalose profile, it segregates from all the other fungi analysed (Figure 3.7). In contrast, if *A. oryzae* NRRL 5590 is removed from the *A. flavus* group for discriminant analysis of the polyol-trehalose profile, it remains grouped with *A. flavus* (data not shown). In addition, if NRRL 274 is removed from the *A. terreus* group for discriminant analysis of the larger subset of GC/MS profile data (Table 3.1), it segregates from the other *A. terreus* reference strains as well as the other Aspergilli (Figure 3.8). It was unclear, based on GC/MS analysis, whether this strain was misidentified based on morphology or was simply an *A. terreus* strain with an atypical biochemical profile. This warranted re-evaluating the identity of NRRL 274 strain as *A. terreus* by additional techniques based on sequence analysis (NRRL 274 is a clinical isolate identified by phenotype). Consequently, the ITS sequence of NRRL 255, 1913, and 274 were amplified by PCR and sequenced to confirm the identity of NRRL 274 as *A. terreus*. The ITS sequences of all three strains were consistent with *A. terreus* sequences in the NCBI database (Figure 3.9).

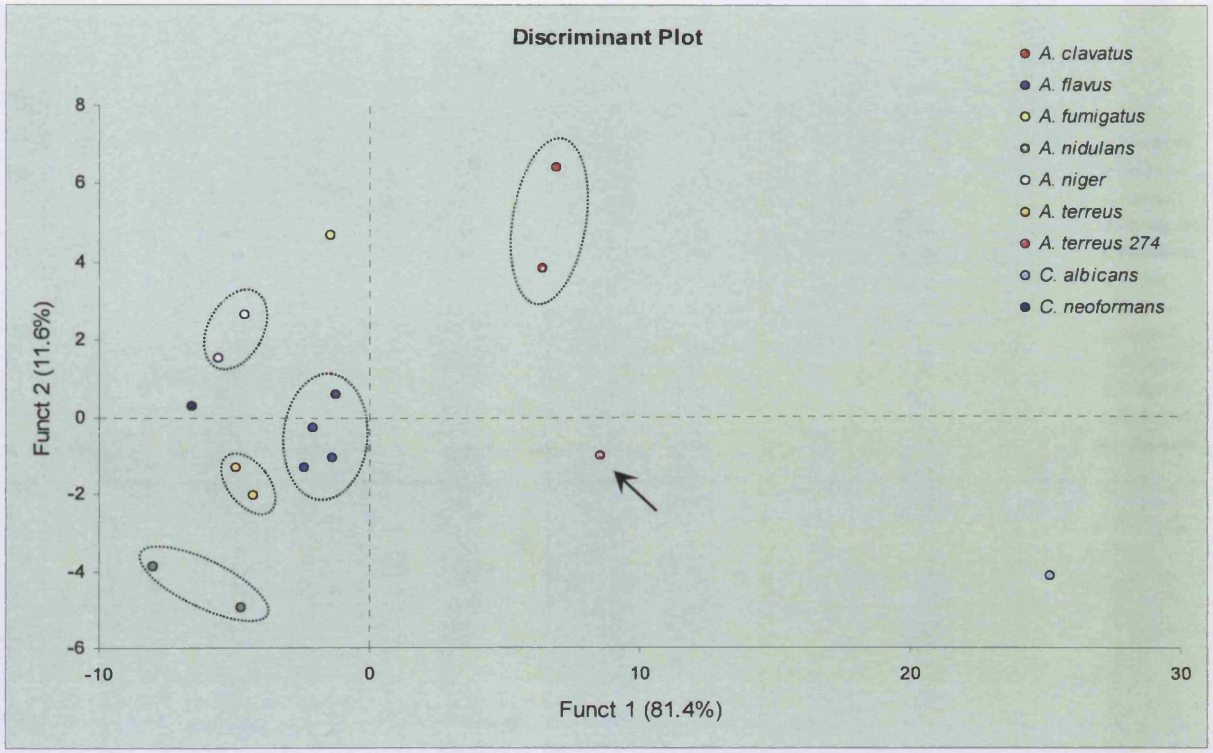


Figure 3.7. Discriminant function analysis based on GC/MS analysis of polyols and trehalose in fungal biomass (NRRL 274 removed from *A. terreus* group). *A. terreus* strain NRRL 274 (data point marked with an arrow) has an unusual polyol-trehalose profile that segregates the strain from the other reference strains analysed. Discrimination between all species improves with removal of NRRL 274 from the *A. terreus* group.

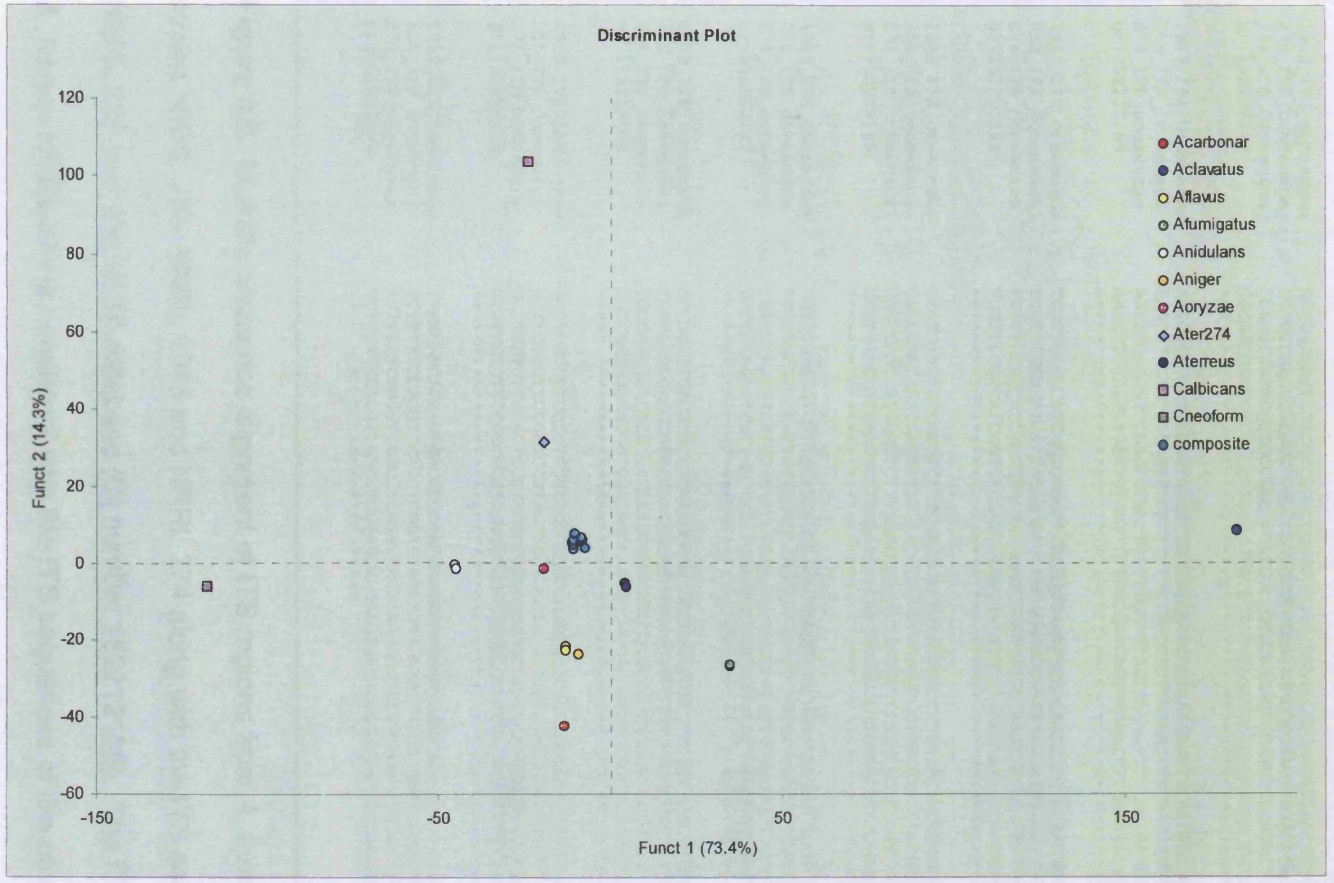
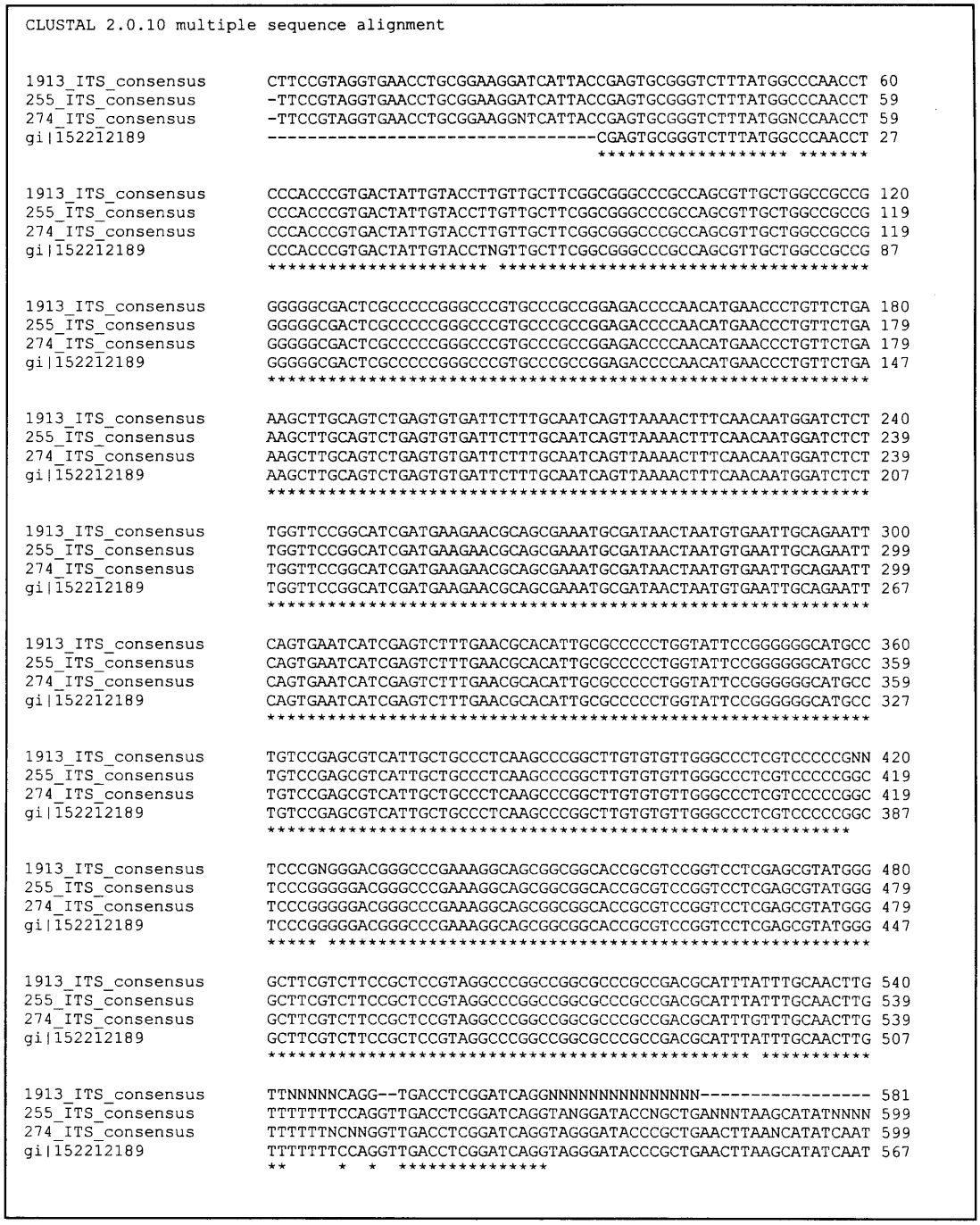


Figure 3.8. Discrimant function analysis of the GC/MS profiling data set (21 compounds) for 17 reference strains and composite samples with NRRL 274 removed from the *A. terreus* group. *A. terreus* NRRL 274 has a polyol distribution that would cause it to segregate from other A. terreus strains and other Aspergilli if labelled as an unknown species.



```
CLUSTAL 2.0.10 multiple sequence alignment


1913_ITS_consensus      CTTCCGTAGGTGAACCTGCGGAAGGATCATTACCGAGTGCGGGTCTTTATGGCCCAACCT 60
255_ITS_consensus       -TTCCGTAGGTGAACCTGCGGAAGGATCATTACCGAGTGCGGGTCTTTATGGCCCAACCT 59
274_ITS_consensus       -TTCCGTAGGTGAACCTGCGGAAGGNTCATTACCGAGTGCGGGTCTTTATGGNCCAACCT 59
gi|152212189            ---------------------------------CGAGTGCGGGTCTTTATGGCCCAACCT 27
                                                 ****************** *******

1913_ITS_consensus      CCCACCCGTGACTATTGTACCTTGTTGCTTCGGCGGGCCCGCCAGCGTTGCTGGCCGCCG 120
255_ITS_consensus       CCCACCCGTGACTATTGTACCTTGTTGCTTCGGCGGGCCCGCCAGCGTTGCTGGCCGCCG 119
274_ITS_consensus       CCCACCCGTGACTATTGTACCTTGTTGCTTCGGCGGGCCCGCCAGCGTTGCTGGCCGCCG 119
gi|152212189            CCCACCCGTGACTATTGTACCTNGTTGCTTCGGCGGGCCCGCCAGCGTTGCTGGCCGCCG 87
                        ********************** ************************************

1913_ITS_consensus      GGGGGCGACTCGCCCCCGGGCCCGTGCCCGCCGGAGACCCCAACATGAACCCTGTTCTGA 180
255_ITS_consensus       GGGGGCGACTCGCCCCCGGGCCCGTGCCCGCCGGAGACCCCAACATGAACCCTGTTCTGA 179
274_ITS_consensus       GGGGGCGACTCGCCCCCGGGCCCGTGCCCGCCGGAGACCCCAACATGAACCCTGTTCTGA 179
gi|152212189            GGGGGCGACTCGCCCCCGGGCCCGTGCCCGCCGGAGACCCCAACATGAACCCTGTTCTGA 147
                        ************************************************************

1913_ITS_consensus      AAGCTTGCAGTCTGAGTGTGATTCTTTGCAATCAGTTAAAACTTTCAACAATGGATCTCT 240
255_ITS_consensus       AAGCTTGCAGTCTGAGTGTGATTCTTTGCAATCAGTTAAAACTTTCAACAATGGATCTCT 239
274_ITS_consensus       AAGCTTGCAGTCTGAGTGTGATTCTTTGCAATCAGTTAAAACTTTCAACAATGGATCTCT 239
gi|152212189            AAGCTTGCAGTCTGAGTGTGATTCTTTGCAATCAGTTAAAACTTTCAACAATGGATCTCT 207
                        ************************************************************

1913_ITS_consensus      TGGTTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAATGTGAATTGCAGAATT 300
255_ITS_consensus       TGGTTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAATGTGAATTGCAGAATT 299
274_ITS_consensus       TGGTTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAATGTGAATTGCAGAATT 299
gi|152212189            TGGTTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAATGTGAATTGCAGAATT 267
                        ************************************************************

1913_ITS_consensus      CAGTGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCC 360
255_ITS_consensus       CAGTGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCC 359
274_ITS_consensus       CAGTGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCC 359
gi|152212189            CAGTGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCC 327
                        ************************************************************

1913_ITS_consensus      TGTCCGAGCGTCATTGCTGCCCTCAAGCCCGGCTTGTGTGTTGGGCCCTCGTCCCCCGNN 420
255_ITS_consensus       TGTCCGAGCGTCATTGCTGCCCTCAAGCCCGGCTTGTGTGTTGGGCCCTCGTCCCCCGGC 419
274_ITS_consensus       TGTCCGAGCGTCATTGCTGCCCTCAAGCCCGGCTTGTGTGTTGGGCCCTCGTCCCCCGGC 419
gi|152212189            TGTCCGAGCGTCATTGCTGCCCTCAAGCCCGGCTTGTGTGTTGGGCCCTCGTCCCCCGGC 387
                        ************************************************************

1913_ITS_consensus      TCCCGNGGGACGGGCCCGAAAGGCAGCGGCGGCACCGCGTCCGGTCCTCGAGCGTATGGG 480
255_ITS_consensus       TCCCGGGGGACGGGCCCGAAAGGCAGCGGCGGCACCGCGTCCGGTCCTCGAGCGTATGGG 479
274_ITS_consensus       TCCCGGGGGACGGGCCCGAAAGGCAGCGGCGGCACCGCGTCCGGTCCTCGAGCGTATGGG 479
gi|152212189            TCCCGGGGGACGGGCCCGAAAGGCAGCGGCGGCACCGCGTCCGGTCCTCGAGCGTATGGG 447
                        ***** ********************************************************

1913_ITS_consensus      GCTTCGTCTTCCGCTCCGTAGGCCCGGCCGGCGCCCGCCGACGCATTTATTTGCAACTTG 540
255_ITS_consensus       GCTTCGTCTTCCGCTCCGTAGGCCCGGCCGGCGCCCGCCGACGCATTTATTTGCAACTTG 539
274_ITS_consensus       GCTTCGTCTTCCGCTCCGTAGGCCCGGCCGGCGCCCGCCGACGCATTTGTTTGCAACTTG 539
gi|152212189            GCTTCGTCTTCCGCTCCGTAGGCCCGGCCGGCGCCCGCCGACGCATTTATTTGCAACTTG 507
                        ********************************************** ************

1913_ITS_consensus      TTNNNNNCAGG--TGACCTCGGATCAGGNNNNNNNNNNNNNNN---------------- 581
255_ITS_consensus       TTTTTTTCCAGGTTGACCTCGGATCAGGTANGGATACCNGCTGANNNTAAGCATATNNNN 599
274_ITS_consensus       TTTTTTNCNNGGTTGACCTCGGATCAGGTAGGGATACCCGCTGAACTTAANCATATCAAT 599
gi|152212189            TTTTTTTCCAGGTTGACCTCGGATCAGGTAGGGATACCCGCTGAACTTAAGCATATCAAT 567
                        **     *  *  ***************
```

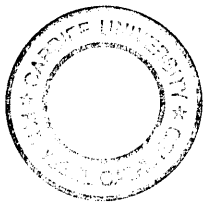
Figure 3.9. Multiple sequence alignment of ITS regions from *A. terreus* reference strains NRRL 255, NRRL 1913 and NRRL 274 along with the ITS sequence of NRRL 255 from the NCBI database (GI number 152212189). The ITS sequence of *A. terreus* NRRL 274 is consistent with the ITS sequences of the other two reference strains and ITS sequences for *A. terreus* on file at NCBI.

Polyols are abundant in fungal samples and the detection and quantitation of these compounds is relatively simple. Because polyols are at such high levels, they may be of value for the general detection of fungi in sample matrices that do not contain significant background levels of polyols. However, identification based on this limited set of compounds (glycerol, erythritol, arabitol, and mannitol) probably would not be sufficient to reliably identify a fungus to species level because these compounds are widely distributed in fungi and a much wider sampling of different species and strains would be required to determine if polyols would be of value as a primary technique for classification of an unknown organism to the Aspergilli. Polyols may be of value as supporting data for speciation when other data, such as morphology and sequence, are available.

The preliminary data from this exploratory study indicated that clustering and discriminant analysis of TMS-derivatized small molecules by GC/MS may be of some use for classifying or characterizing unknown species in larger taxonomic studies, but to confirm the utility of this approach, the experiment would need to be replicated with the same reference set and additional sets of test samples. Also, this approach would be impractical for routine detection or identification of specific fungi, because it requires concurrent analysis of a large set of reference samples because although the profiles obtained should be similar under controlled culture conditions, the biochemical profiles will be influenced by small variations in experimental conditions. Biochemical profiling requires relatively expensive equipment as well as skilled operators and the resulting data files are quite large and require considerable computer resources and human input for processing. Consequently, the time and resources required to identify fungi with this approach are excessive when compared to other more commonly used techniques.

Fatty Acid Methyl Ester and Neutral Lipid Analyses. Solvent extracts of mycelial homogenates were analyzed by GC/MS to determine fatty acid methyl ester (FAME) and neutral lipid profiles for the test samples. The analyses produced data on a very limited number of C16 and C18 fatty acids, ergosterol and related sterols, along with some minor unidentified components. Even under controlled culture conditions, the fatty acid profiles were relatively uninformative and exhibited a great deal of variation within species and were of little practical use for classifying the fungi. No unique compounds were detected that would allow speciation of pure cultures or detection of a specific genus of fungus within a complex biological sample. As with polyols, the presence of ergosterol in a lipid extract may be of some value for general detection of fungi in sample matrices that do not contain significant background levels of ergosterol.

During method development, significant levels farnesol, methyl farnesoate, and related isoprenoid compounds were readily detected in lipid extracts of two strains of *A. nidulans* (A4 and NRRL 187). These were not detected in other species, despite concentrating samples to an equivalent of 400 mg fresh weight mycelium per milliliter of final extract (Figure 3.10). Farnesol and methyl farnesoate are of interest because there is a growing body of evidence that these compounds have important roles in



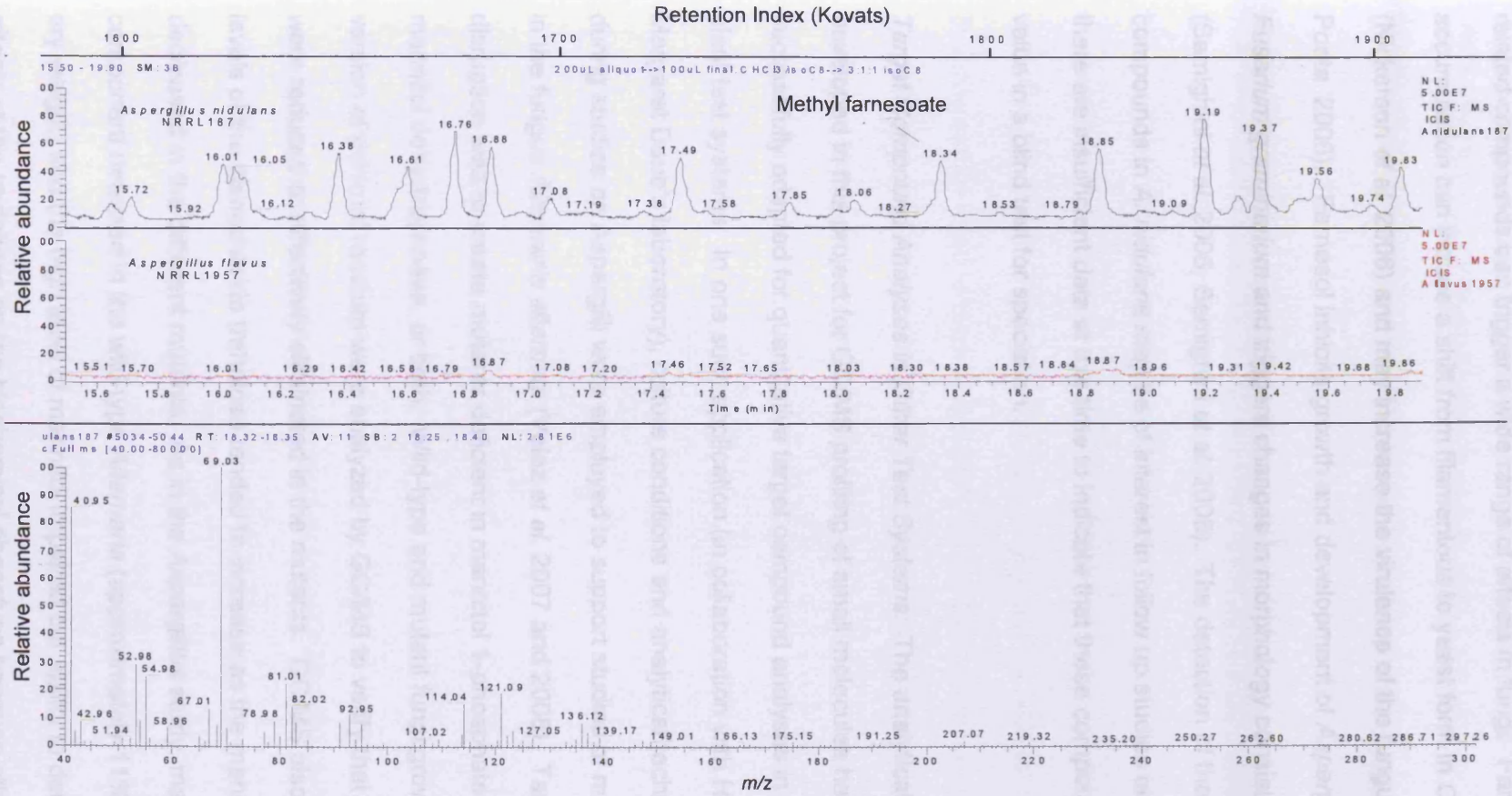


Figure 3.10. GC/MS chromatograms of lipid extracts from *Aspergillus nidulans* and *Aspergillus flavus* (top) showing series of chromatographic peaks for methyl farnesoate and related compounds in *A. nidulans*. Chromatograms for *A. nidulans* and *A. flavus* are on same scale and are equivalent to the same amount of starting material. Methyl farnesoate and related compounds are absent from the *A. flavus* extract. Electron impact (EI) spectrum of methyl farnesoate is shown (bottom).

the growth and development of a number of different organisms. Insect juvenile hormones are derivatives of farnesoic acid (Holstein and Hohl 2004). Farnesol and related compounds can trigger a wide range of effects in fungi. Farnesol accumulation can induce a shift from filamentous to yeast form in *C. albicans* (Nickerson *et al.* 2006) and may increase the virulence of the fungus (Shea and Del Poeta 2006). Farnesol inhibits growth and development of *Aspergillus nidulans* and *Fusarium graminarium* and triggers changes in morphology consistent with apoptosis (Semighini *et al.* 2006, Semighini *et al.* 2008). The detection of these isoprenoid compounds in *A. nidulans* may be of interest in follow up studies on metabolism, but there are insufficient data at this time to indicate that these compounds would be of value in a blind test for speciation.

Target Compound Analyses in Other Test Systems. The analytical methodology developed in this project for GC/MS profiling of small molecules has been successfully adapted for quantitative target compound analyses in other fungal and plant test systems. In one such application (in collaboration with Heriberto Vélez in Margaret Daub's laboratory), culture conditions and analytical techniques developed during studies on Aspergilli were employed to support studies of mannitol synthesis in the fungus *Alternaria alternata* (Velez *et al.* 2007 and 2008). Targeted gene disruption was to create mutants deficient in mannitol 1-phosphate 5-dehydrogenase, mannitol dehydrogenase, or both. Wild-type and mutant fungi grown in an early version of minimal medium were analyzed by GC/MS to verify that mannitol levels were reduced or effectively eliminated in the mutants. GC/MS also revealed that levels of the disaccharide trehalose tended to increase as the mannitol levels decreased in the different mutants. As in the *Aspergillus* study, mannitol was a major component detected in the wild-type *Alternaria* (approximately 11% of the mycelial dry weight) and the high level of mannitol impaired our ability to determine pleiotropic effects of the mutations on the biochemical phenotype because other components of

potential interest were at such low levels relative to the mannitol in the samples. Targeted analyses of other compounds using alternative methods on replicate samples was precluded because some of the mutants grew very slowly and only very limited amounts were available for some samples.

Summary. In this study, the objective was to find compounds that may be unique to certain species of *Aspergillus* or to find patterns in the small molecule phenotypes for a cross section of *Aspergillus* species that would allow classification of unknown strains. Small molecule profiling by GC/MS was useful as a survey of the overall composition of test systems and the methodology provided a good basis for developing targeted analyses. Profiling data provided by GC/MS analyses of fungal biomass can be used to classify Aspergilli, but no clear patterns emerged from specific classes of chemicals that would allow routine classification of unknown organisms without the analysis of a relatively large number of reference samples to verify accurate groupings.

Chapter 4: Results of LC/MS Analyses

Liquid chromatography coupled with mass spectrometry (LC/MS) was used in this study to conduct targeted analyses, as well as small molecule and protein profiling of fungal samples to look for compounds or peptide sequences unique to different species of Aspergilli. Targeted analyses were directed at mycotoxins, such as aflatoxins and cyclopiazonic acid, that are produced by a relatively small group of Aspergilli. Detection of these compounds can confirm the identity of an organism as one of the aflatoxin producing species or exclude classification as one of the non-aflatoxin producers. Small molecule profiling by LC/MS, as with GC/MS, was conducted to catalogue a cross section of small molecules produced by Aspergilli and look for specific compounds or patterns within groups of compounds that might be used to classify the test strains to species level. Protein analysis was initially conducted in an attempt to identify a macromolecule detected during small molecule profiling, but developed into a profiling technique that permitted identification of organisms directly from the characteristic amino acid sequences of peptides from tryptic digests of total protein extracts.

Mycotoxin analysis. LC/MS performed well for the quantitative targeted analysis of aflatoxins in the range of approximately 10ng on column down to 10pg on column. A representative chromatogram and example spectra of a mixed standard is presented in Figure 4.1. Aflatoxin analysis can be used as an aid to identification of some *Aspergillus* species, but only as confirmation because aflatoxins are best produced on specialized media and are not consistently produced by all isolates of the species known to produce aflatoxins - *A. flavus, A. parasiticus, A. nomius*, *A. pseudotamarii*, and *A. bombycis* (Scheidegger and Payne, 2003). When grown in liquid culture with BCP minimal medium, aflatoxin production was relatively low (several hundred ng/mL) and variable for strains known to produce aflatoxins and consequently was of little use

as a primary characteristic for classification of Aspergilli. Some target compound analyses (cyclopiazonic acid, norsolorinic acid, versacolorin A) required reconfiguring the LC/MS system with an alkaline mobile phase that interfered with other ongoing analyses and these analyses were not pursued further for this project.

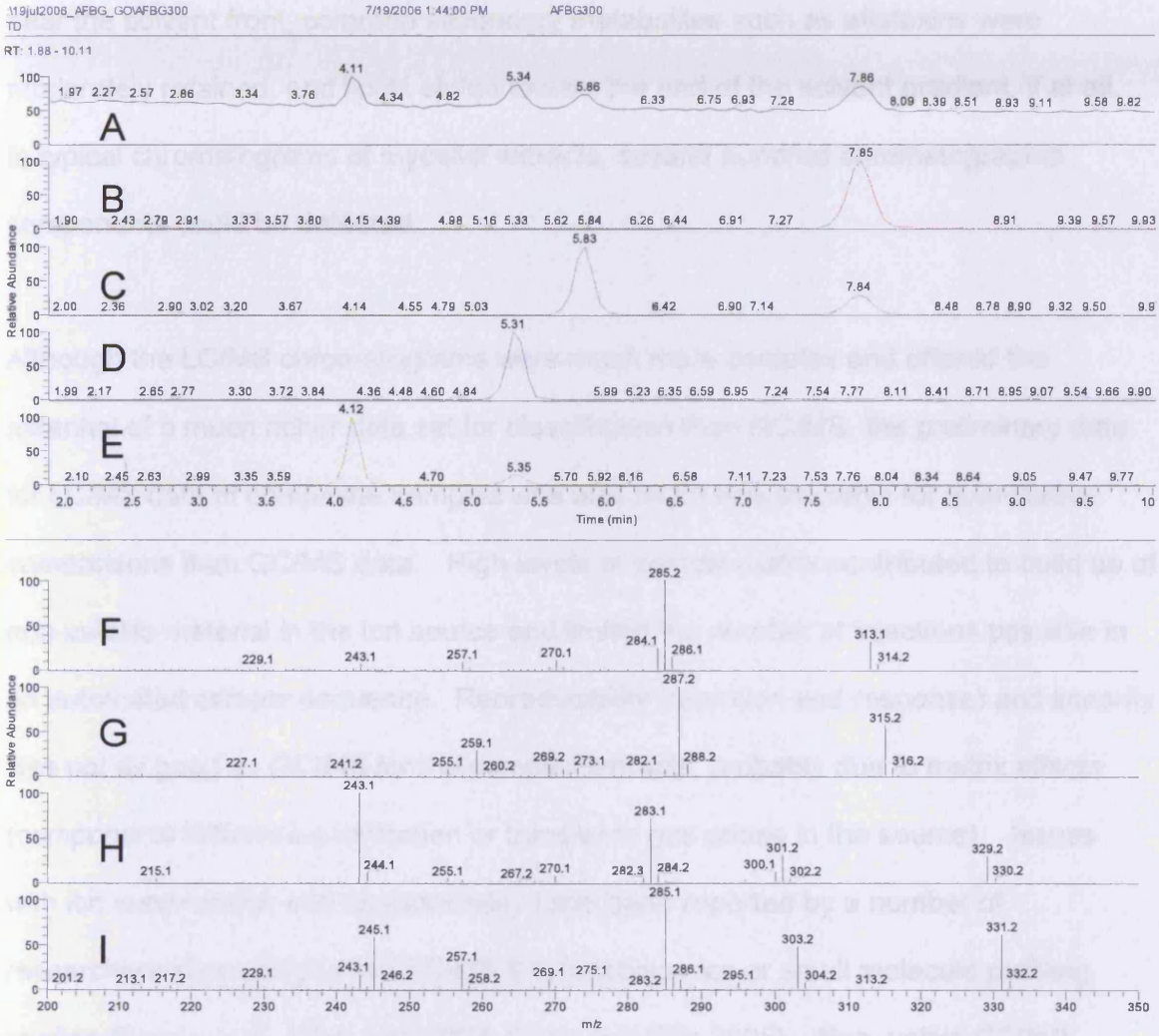


Figure 4.1. Typical LC/MS/MS chromatogram of a mixed aflatoxin reference standard, 300ng/mL (A-E) and spectra for aflatoxins (F-I). Chromatographic traces - (A) Total ion chromatogram (TIC); (B) extracted ion chromatogram (XIC) aflatoxin B1, *m/z* 313; (C) XIC aflatoxin B2, *m/z* 315; (D) XIC aflatoxin G1, *m/z* 329; (E) XIC aflatoxin G2, *m/z* 331. LC/MS/MS spectra - (F) aflatoxin B1, (G) aflatoxin B2, (H) aflatoxin G1, and (I) aflatoxin G2.

Small Molecule Profiling by LC/MS. Profiling of small molecules yielded much more complex chromatograms than those for GC/MS profiling. An example of a chromatogram for a composite sample is presented in Figure 4.2. The separation for LC/MS profiling employed reverse phase chromatography, so highly polar compounds such as amino acids were relatively unretained on the C18 column and eluted at or near the solvent front, common secondary metabolites such as aflatoxins were moderately retained, and lipids eluted toward the end of the solvent gradient, if at all. In typical chromatograms of mycelial extracts, several hundred chromatographic components could be detected.

Although the LC/MS chromatograms were much more complex and offered the potential of a much richer data set for classification than GC/MS, the preliminary data for LC/MS data of composite samples was also much less tractable for quantitative comparisons than GC/MS data. High levels of sample matrix contributed to build up of non-volatile material in the ion source and limited the number of injections possible in an automated sample sequence. Reproducibility (retention and response) and linearity was not as good as GC/MS for the complex extracts, probably due to matrix effects (components influencing ionization or transfer to gas phase in the source). Issues with ion suppression and reproducibility have been reported by a number of researchers attempting to use LC/MS for metabolomics or small molecule profiling studies (Fernie *et al.* 2004, Kell 2004, Dunn and Ellis 2005). Also, unlike GC/MS analysis, there are no large spectral databases produced under standard conditions for identification of small molecules from LC/MS or LC/MS/MS spectra, so identification of the chromatographic peaks of interest was very difficult. Consequently, the methodology for LC/MS profiling would have required significantly more development to refine the technique for quantitative comparisons. This method development, in part, would have involved the use of smaller bore columns and longer solvent

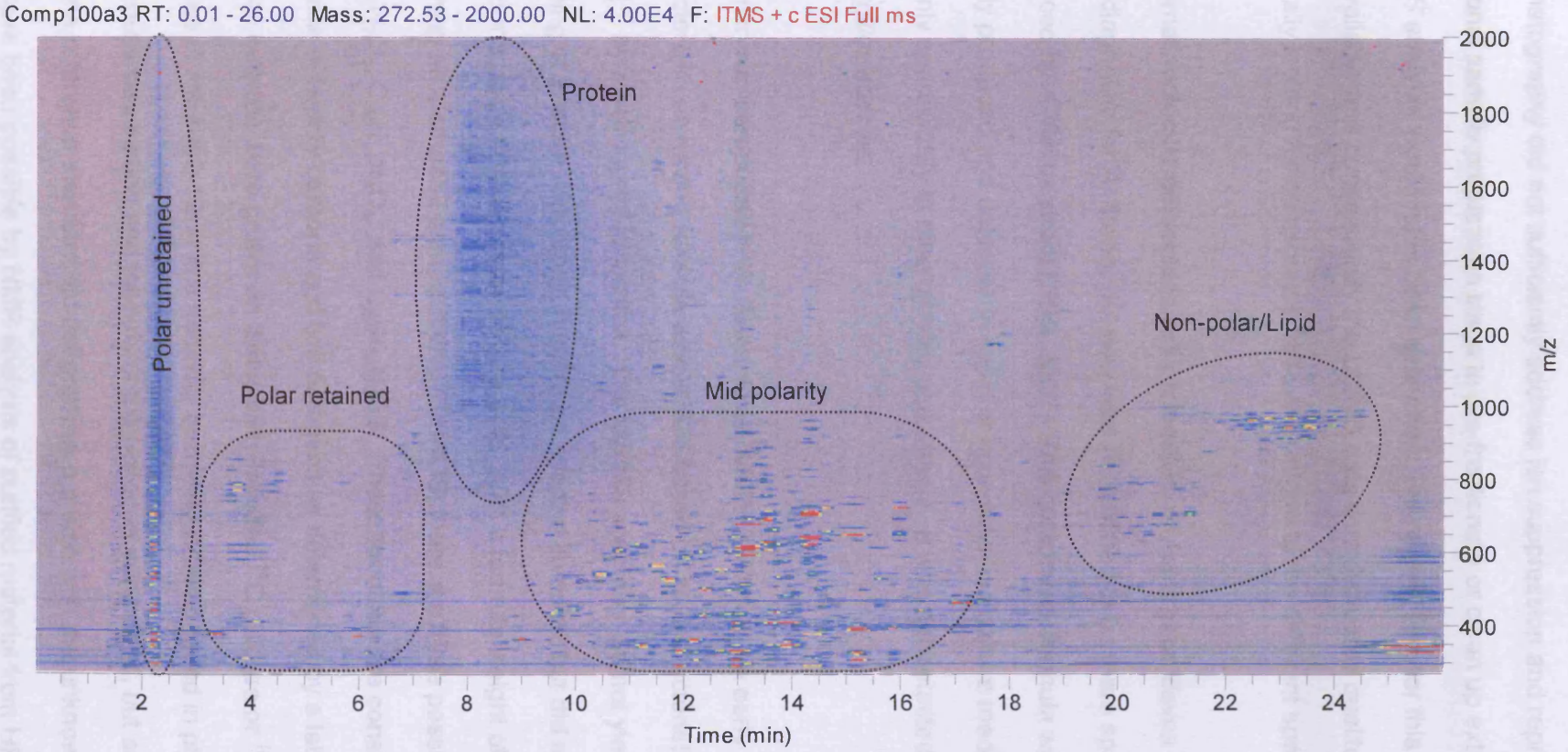


Figure 4.2. Typical LC/MS chromatogram for small molecule profiling of a composite sample, 100 mg fresh weight per millilitre, prepared from fungal biomass homogenates of reference strains. The chromatogram is presented as a 3D plot with time on the x-axis, *m/z* on the y-axis, and intensity on the z-axis indicated by colour (grey background, blue least intense, red most intense).

gradients to achieve greater sensitivity and better chromatographic resolution while loading a smaller amount of matrix in each injection. If improvements in chromatography did not sufficiently address ion suppression and reproducibility, additional sample preparation steps to pre-fractionate or clean up extracts prior to LC/MS analysis would have been evaluated. Instrument time for this development was not available and consequently the LC/MS data was evaluated qualitatively by manually looking for compounds that were unique to the different species.

One small molecule detected in culture medium of *Aspergillus flavus* initially appeared to be diagnostic for *A. flavus*, but was later determined to be more specifically produced by *A. flavus* strain NRRL 3357. This compound, formula weight 326, was reliably produced and detected in liquid or agar-solidified culture medium of this strain and only sporadically in other strains of *A. flavus*. It was not detected in other *Aspergillus* species.

The unknown compound in *A. flavus* strain NRRL 3357 was an early eluting compound characterized as having formula weight 326 based on its characteristic ion *m/z* 327 $[M+H]^{+1}$ by electrospray ionization. The MS/MS spectrum did not yield identification by matching of spectra with those in common spectral libraries, but did indicate that the compound had a ring structure (Figure 4.3). With a formula weight of 326 and assuming an elemental composition of CHNOS, there are three possible formulae – $C_{20}H_{26}N_2O_2$, $C_{20}H_{10}N_2O_3$, and $C_{20}H_{10}N_2OS$. These formulae are consistent with the presence of twenty carbons and two nitrogens as determined by a labelling experiment in which cultures were grown on isotopically-labelled ^{13}C glucose or ^{15}N ammonia. Ajmaline (CAS 4360-12-7) is a naturally occurring alkaloid found in plants that matched the general description and formula of the unknown compound, but analysis of an ajmaline reference standard did not produce a match with the unknown. Identification may have been possible by NMR analysis of purified material from HPLC fractionation,

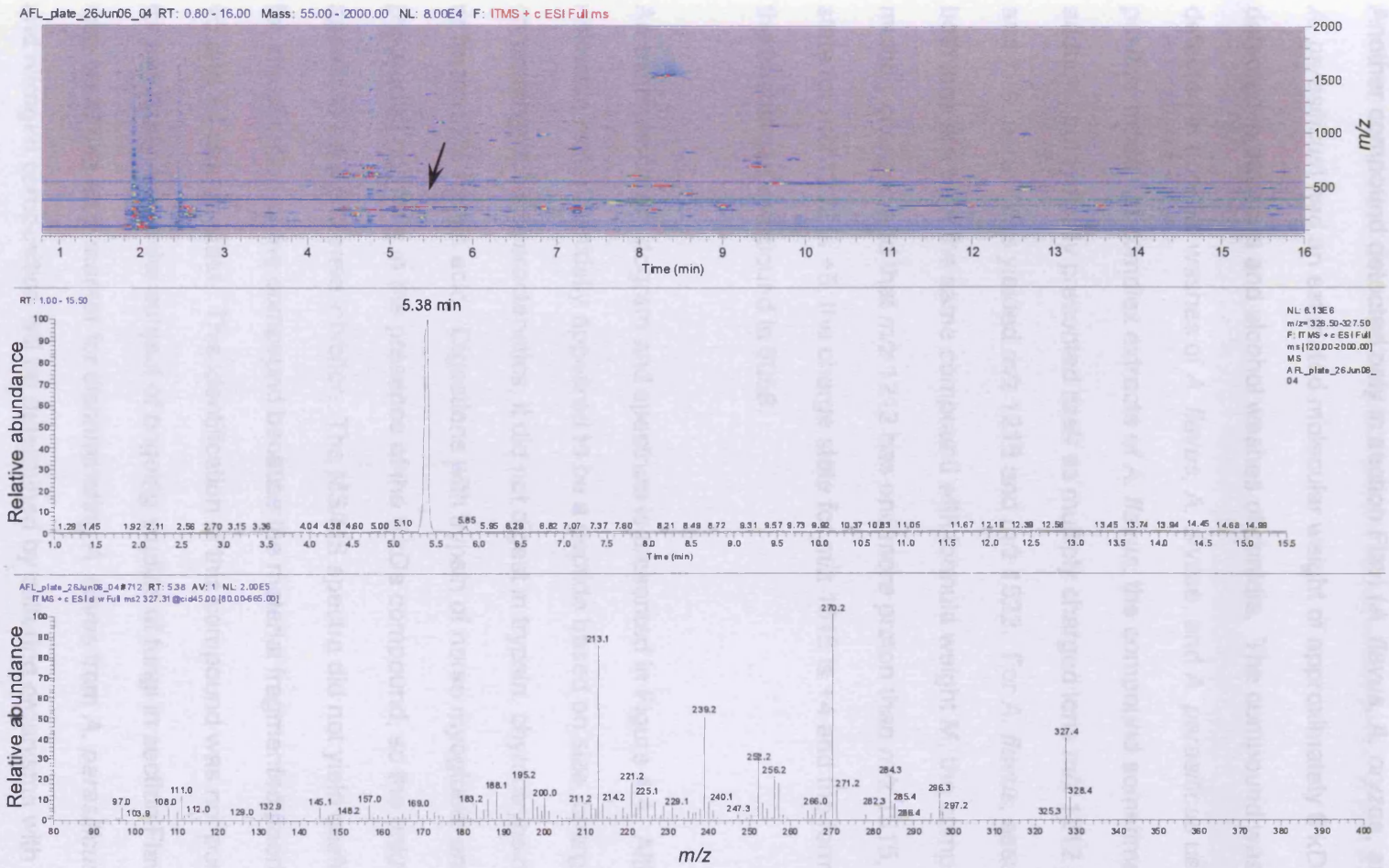


Figure 4.3. Typical LC/MS/MS chromatogram and spectrum of an unknown small molecule from *A. flavus* medium - (A) Total ion chromatogram (TIC), 3D plot, chromatographic peak of unknown marked with arrow; (B) extracted ion chromatogram (XIC) unknown compound, *m/z* 327 $[M+H]^{+1}$; (C) MS/MS spectrum of unknown compound.

but due to the limited value of this compound for speciation of *Aspergillus* cultures, identification of this compound was not pursued further.

Another compound detected only in section Flavi (*A. flavus*, *A. oryza*e, and *A. parasiticus*) had an estimated molecular weight of approximately 6 kDa and was detected in aqueous and alcohol washes of conidia. The compound was reliably detected in conidial washes of *A. flavus*, *A. oryza*e, and *A. parasiticus* using ESI in positive mode. In complex extracts of *A. flavus*, the compound sometimes formed adducts, but typically presented itself as multiply charged ions *m/z* 1212 and *m/z* 1515 and in *A. parasiticus* yielded *m/z* 1218 and *m/z* 1522. For *A. flavus*, assuming that both ions are from the same compound with formula weight M, the compound is multiply charged and that *m/z* 1212 has one more proton than *m/z* 1515, the charge state for *m/z* 1212 is +5, the charge state for m/z 1515 is +4 and the formula weight of the uncharged compound is 6056.

An example chromatogram and spectrum is presented in Figure 4.4. Although this unknown material initially appeared to be a peptide based on size, charge, and chromatographic characteristics, it did not digest in trypsin, chymotrypsin, Glu C (V8), or heated 2% formic acid. Digestions with trypsin of horse myoglobin and BSA proceeded normally in the presence of the 6 kDa compound, so the unknown did not appear to be a protease inhibitor. The MS/MS spectra did not yield useful information for identification of the compound because the material fragmented poorly and retained +5 and +4 charge state. This identification of this compound was not pursued further for this study, but is the subject of ongoing studies of fungi in section Flavi because it may be of use as a marker for discriminating *A. flavus* from *A. parasiticus*. The carbon and nitrogen composition will be determined by culturing organisms with isotopically-labelled ^{13}C glucose or ^{15}N ammonia, in the same manner as the unknown compound

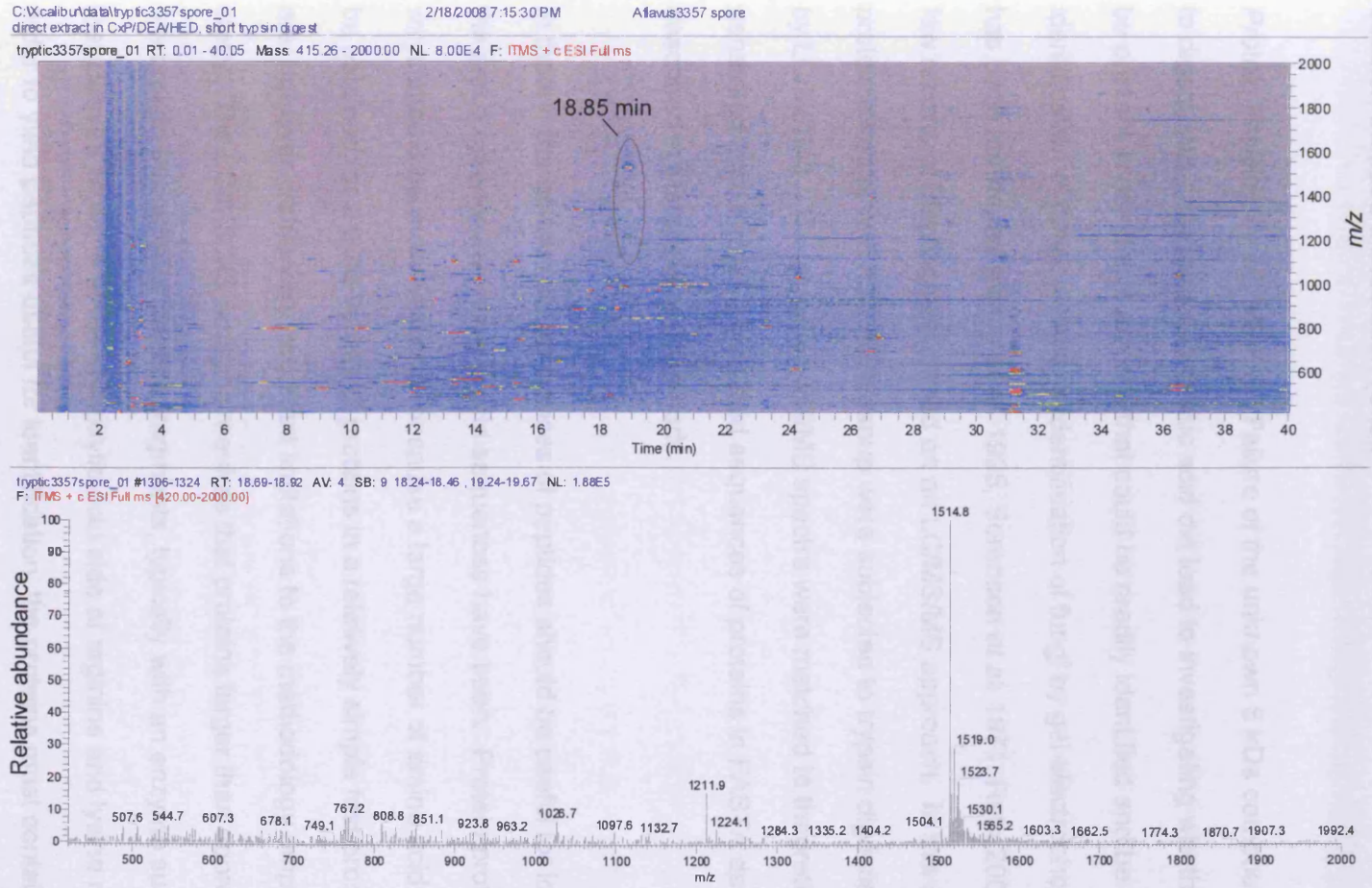


Figure 4.4. Typical LC/MS/MS chromatogram and spectrum of unknown 6kDa compound from *A. flavus* conidial washes after trypsin digest - (A) Total ion chromatogram (TIC), 3D plot, chromatographic peak of unknown marked with oval; (B) LC/MS spectrum of unknown compound, multiply charged ions *m/z* 1212 $[M+H]^{+5}$ and *m/z* 1515 $[M+4H]^{+4}$ shown. Unknown compound is present before and after trypsin digestion.

of molecular weight 326 from *A. flavus* strain NRRL 3357. In addition, the compound will be characterized by LC/MS/MS equipped with different types of mass spectrometers with greater mass accuracy and different types of ion fragmentation, such as electron-transfer dissociation (ETD) and high-energy collisional dissociation (HCD) that may yield more useful MS/MS spectra than were obtained in this study by collision-induced dissociation (CID).

Protein Profiling by LC/MS/MS. Failure of the unknown 6 kDa compound from conidia to digest with proteases and formic acid did lead to investigating whether proteins may be present in conidia or mycelia that could be readily identified and be useful for identification of fungi. Although identification of fungi by gel electrophoresis of proteins has been attempted (Rinyu *et al.* 1995, Sorenson *et al.* 1971, Rath 2001), there are few reports of identification based on an LC/MS/MS approach. In this study, total protein extracts with minimal cleanup were subjected to trypsin digestion and analyzed by LC/MS/MS. The resulting MS/MS spectra were matched to theoretical spectra generated from known amino acid sequences of proteins in FASTA databases as described in Materials and Methods.

In theory, the amino acid sequences of peptides should be useful for identification in the same manner that nucleic acid sequences have been. Protein profiling in some ways should be more versatile, because a large number of amino acid sequences can be obtained for a wide variety of proteins in a relatively simple fast procedure. There are, however, some very important limitations to the methodology employed in this study. The LC/MS/MS analysis requires that proteins larger than approximately 6 kDa are broken into smaller peptide fragments, typically with an enzyme such as trypsin that cleaves proteins at the carboxylic acid side of arginine and lysine residues. In order to yield peptides useful for identification, the proteins must contain arginine and lysine residues in positions such that the resulting peptides are in the range of 5-60

amino acid residues and molecular weights in the range of 400-6000 Da. In the mass spectrometer, the resulting peptides must yield positive ions with charge states of 1-3 and fragment by MS/MS to produce spectra suitable for matching with theoretical spectra generated from a FASTA database. Proteins useful for classifying fungi would have taxonomically unique tryptic peptide sequences that may be predicted by multiple sequence alignment software and can be readily detected by the profiling methodology.

A typical LC/MS/MS analysis of a trypsin digest yields over 1000 spectra and 50-150 identified proteins in a single chromatographic run. Example chromatograms for trypsin digests of *A. flavus* conidia and mycelia are presented in Figures 4.5 and 4.6, respectively. An example of a SEQUEST® search through BioWorks™, matching LC/MS/MS spectra to proteins, is presented in Figure 4.7. In this example, one of the top proteins detected is enolase and the entry has been expanded to show the tryptic peptides from enolase detected in the LC/MS/MS analysis. Also shown, are the calculated probabilities for correct matches to individual peptides and the specific protein. An example of how the MS/MS spectrum is matched to an amino acid sequence of a peptide from enolase is presented in Figure 4.8. A table summarizing the nomenclature for amino acids in proteins and the masses of residues used for interpreting MS/MS spectra of peptides is presented in Table 4.1.

amino acid	abbrev	letter	-NH-CHR-CO- residue	amino acid formula	amino acid ave fw	notes
glycine	Gly	G	57.02	$C_2H_5NO_2$	75.07	Nonpolar
alaninine	Ala	A	71.04	$C_3H_7NO_2$	89.09	Nonpolar
serine	Ser	S	87.03	$C_3H_7NO_3$	105.09	Polar
proline	Pro	P	97.05	$C_5H_9NO_2$	115.13	Nonpolar
valine	Val	V	99.07	$C_5H_{11}NO_2$	117.15	Nonpolar
threonine	Thr	T	101.05	$C_4H_9NO_3$	119.12	Polar
cysteine	Cys	C	103.01	$C_3H_7NO_2S$	121.16	Polar
isoleucine	Ile	I	113.08	$C_6H_{13}NO_2$	131.17	Nonpolar
leucine	Leu	L	113.08	$C_6H_{13}NO_2$	131.17	Nonpolar
asparagine	Asn	N	114.04	$C_4H_8N_2O_3$	132.12	Polar
aspartic acid	Asp	D	115.03	$C_4H_7NO_4$	133.10	Polar, neg
lysine	Lys	K	128.09	$C_6H_{14}N_2O_2$	146.19	Polar, pos
glutamine	Gln	Q	128.06	$C_5H_{10}N_2O_3$	146.15	Polar
glutamic acid	Glu	E	129.04	$C_5H_9NO_4$	147.13	Polar, neg
methionine	Met	M	131.04	$C_5H_{11}NO_2S$	149.21	Nonpolar
histidine	His	H	137.06	$C_6H_9N_3O_2$	155.16	Polar, pos
phenylalanine	Phe	F	147.07	$C_9H_{11}NO_2$	165.19	Nonpolar
arginine	Arg	R	156.10	$C_6H_{14}N_4O_2$	174.20	Polar, pos
tyrosine	Tyr	Y	163.06	$C_9H_{11}NO_3$	181.19	Polar
tryptophan	Trp	W	183.08	$C_{11}H_{12}N_2O_2$	204.23	Nonpolar

Table 4.1. Useful amino acid reference information for reading sequences and interpreting spectra. The mass listed under '-NH-CHR-CO- residue' is the formula weight of the free amino acid minus the water loss from forming peptide bonds.

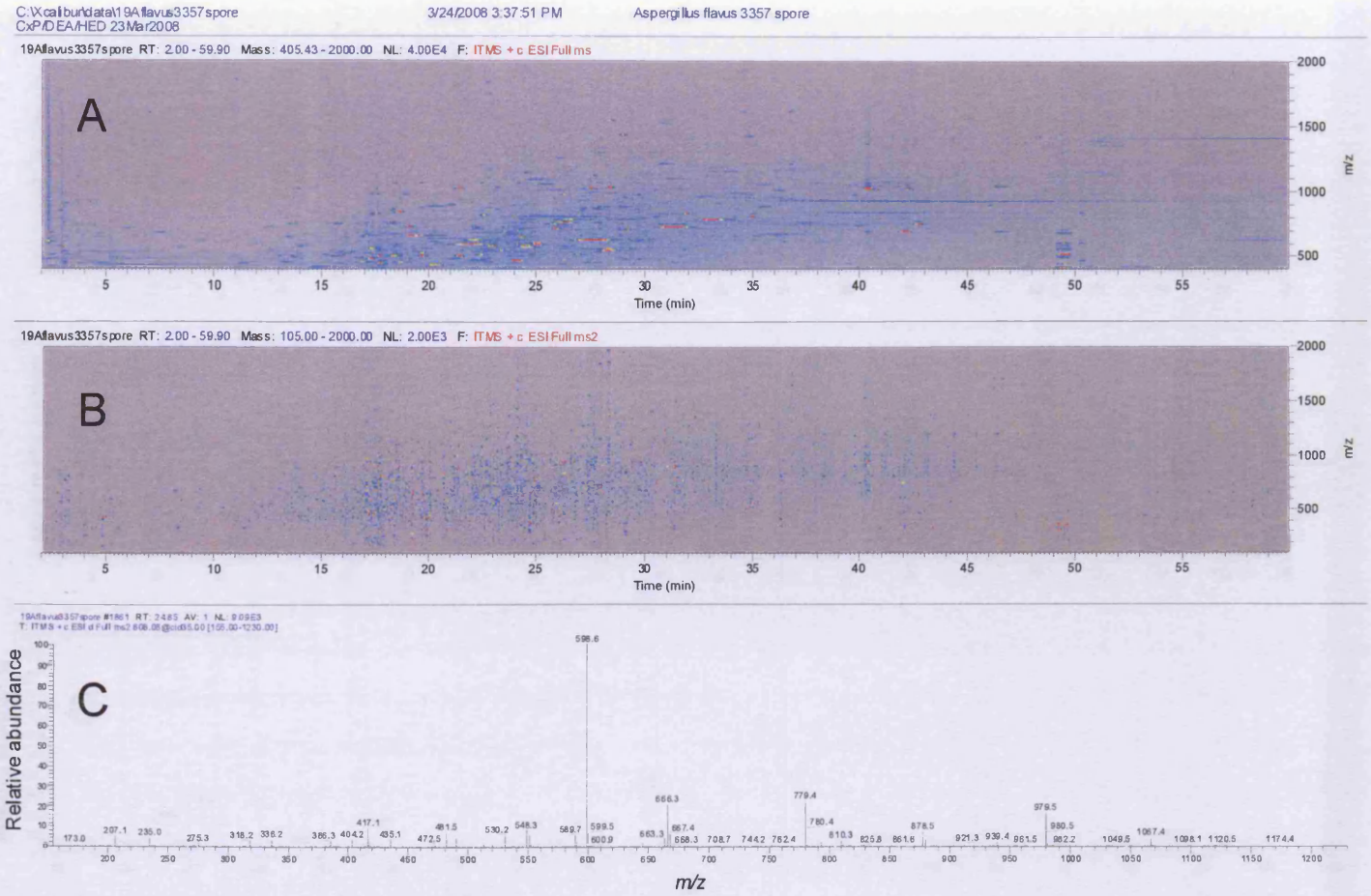


Figure 4.5. Typical LC/MS/MS chromatogram and spectra of trypsin digest of an *A. flavus* conidial protein extract - (A) Total ion chromatogram (TIC), 3D plot; (B) MS/MS spectra of four most intense ions in MS scan, 3D plot; (C) example MS/MS spectrum.

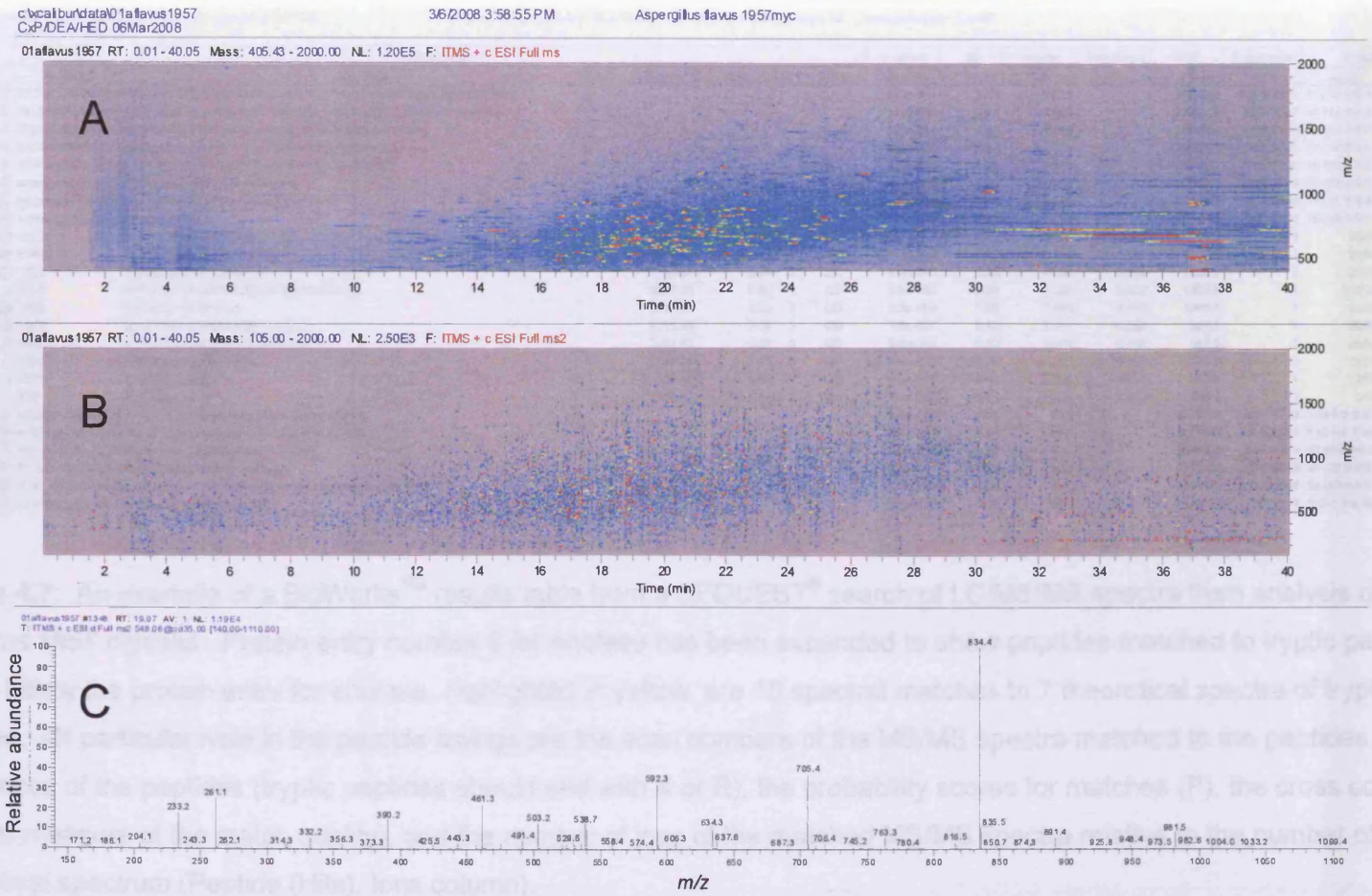
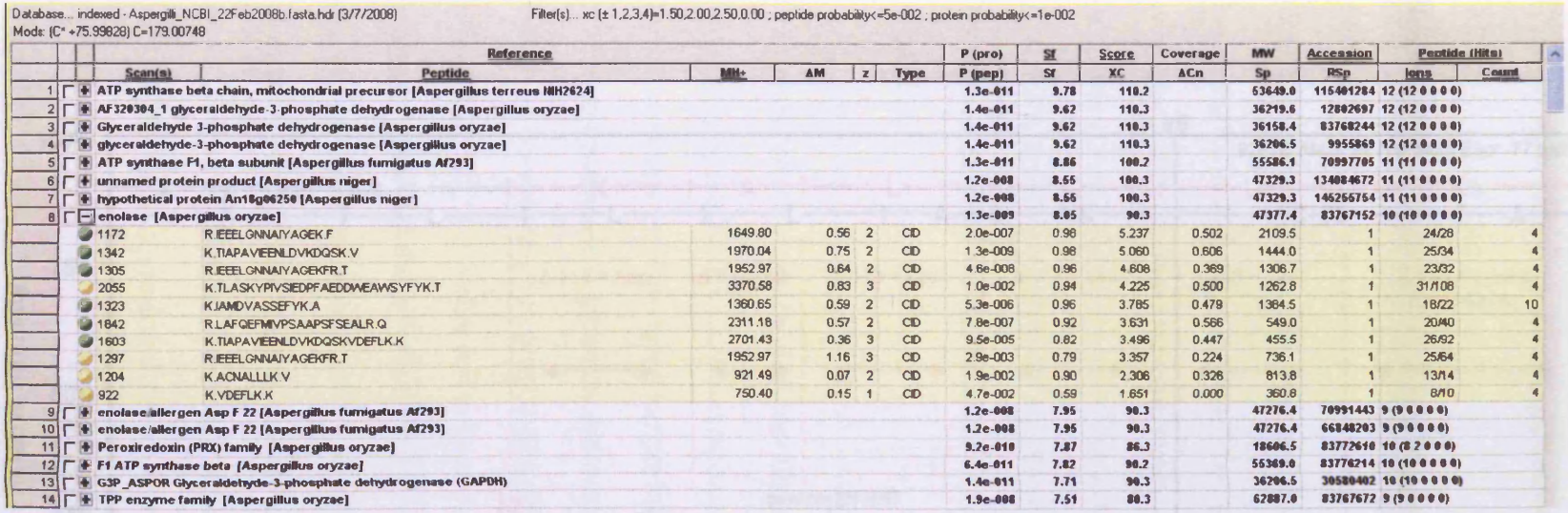


Figure 4.6. Typical LC/MS/MS chromatogram and spectra of trypsin digest of an *A. flavus* mycelial protein extract - (A) Total ion chromatogram (TIC), 3D plot; (B) MS/MS spectra of four most intense ions in MS scan, 3D plot; (C) example MS/MS spectrum.



Database... indexed - Aspergili_NCBI_22Feb2008b.fasta.hdr (3/7/2008) Filter(s)... xc (± 1,2,3,4)=1.50,2.00,2.50,0.00 ; peptide probability<=5e-002 ; protein probability<=1e-002
Mods: (C* +75.99828) C=179.00748

		Reference						P (pro)	Sf	Score	Coverage	MW	Accession	Peptide (Hits)	
	Scan(s)	Peptide	MH+	AM	z	Type	P (pep)	Sf	XC	ACn	Sp	RSp	Ions	Count
1		ATP synthase beta chain, mitochondrial precursor [Aspergillus terreus NIH2624]					1.3e-011	9.78	110.2		53649.0	115401284	12 (12 0 0 0 0)	
2		AF320304_1 glyceraldehyde-3-phosphate dehydrogenase [Aspergillus oryzae]					1.4e-011	9.62	110.3		36219.6	12802697	12 (12 0 0 0 0)	
3		Glyceraldehyde 3-phosphate dehydrogenase [Aspergillus oryzae]					1.4e-011	9.62	110.3		36158.4	83768244	12 (12 0 0 0 0)	
4		glyceraldehyde-3-phosphate dehydrogenase [Aspergillus oryzae]					1.4e-011	9.62	110.3		36206.5	9955869	12 (12 0 0 0 0)	
5		ATP synthase F1, beta subunit [Aspergillus fumigatus Af293]					1.3e-011	8.86	100.2		55586.1	70997705	11 (11 0 0 0 0)	
6		unnamed protein product [Aspergillus niger]					1.2e-008	8.55	100.3		47329.3	134084672	11 (11 0 0 0 0)	
7		hypothetical protein An18g06250 [Aspergillus niger]					1.2e-008	8.55	100.3		47329.3	145256754	11 (11 0 0 0 0)	
8		enolase [Aspergillus oryzae]					1.3e-009	8.05	90.3		47377.4	83767152	10 (10 0 0 0 0)	
	1172	R.IEEELGNNAIYAGEK.F	1649.80	0.56	2	CID	2.0e-007	0.98	5.237	0.502	2109.5	1	24/28	4
	1342	K.TIAPAVIEENLDVKDQSK.V	1970.04	0.75	2	CID	1.3e-009	0.98	5.060	0.606	1444.0	1	25/34	4
	1305	R.IEEELGNNAIYAGEKFR.T	1952.97	0.64	2	CID	4.6e-008	0.96	4.608	0.369	1306.7	1	23/32	4
	2055	K.TLASKYPIVSIEDPFAEDDWEAWSYFYK.T	3370.58	0.83	3	CID	1.0e-002	0.94	4.225	0.500	1262.8	1	31/108	4
	1323	K.IAMDVASSEFYK.A	1360.65	0.59	2	CID	5.3e-006	0.96	3.785	0.479	1364.5	1	18/22	10
	1842	R.LAFQEFMIVPSAAPSFSEALR.Q	2311.16	0.57	2	CID	7.8e-007	0.92	3.631	0.566	549.0	1	20/40	4
	1603	K.TIAPAVIEENLDVKDQSKVDEFLK.K	2701.43	0.36	3	CID	9.5e-005	0.82	3.496	0.447	455.5	1	26/92	4
	1297	R.IEEELGNNAIYAGEKFR.T	1952.97	1.16	3	CID	2.9e-003	0.79	3.357	0.224	736.1	1	25/64	4
	1204	K.ACNALLLK.V	921.49	0.07	2	CID	1.9e-002	0.90	2.306	0.326	813.8	1	13/14	4
	922	K.VDEFLK.K	750.40	0.15	1	CID	4.7e-002	0.59	1.651	0.000	360.8	1	8/10	4
9		enolase/allergen Asp F 22 [Aspergillus fumigatus Af293]					1.2e-008	7.95	90.3		47276.4	70991443	9 (9 0 0 0 0)	
10		enolase/allergen Asp F 22 [Aspergillus fumigatus Af293]					1.2e-008	7.95	90.3		47276.4	66848203	9 (9 0 0 0 0)	
11		Peroxiredoxin (PRX) family [Aspergillus oryzae]					9.2e-010	7.87	86.3		18606.5	83772610	10 (8 2 0 0 0)	
12		F1 ATP synthase beta [Aspergillus oryzae]					6.4e-011	7.82	90.2		55369.0	83776214	10 (10 0 0 0 0)	
13		G3P_ASPOR Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)					1.4e-011	7.71	90.3		36206.5	30580402	10 (10 0 0 0 0)	
14		TPP enzyme family [Aspergillus oryzae]					1.9e-008	7.51	80.3		62887.0	83767672	9 (9 0 0 0 0)	

Figure 4.7. *An example of a BioWorks™ results table from a SEQUEST® search of LC/MS/MS spectra from analysis of tryptic digests of A. flavus 1957 mycelia. Protein entry number 8 for enolase has been expanded to show peptides matched to tryptic peptides of enolase.* Listed below the protein entry for enolase, highlighted in yellow, are 10 spectral matches to 7 theoretical spectra of tryptic peptides from enolase. Of particular note in the peptide listings are the scan numbers of the MS/MS spectra matched to the peptides, the amino acid sequences of the peptides (tryptic peptides should end with K or R), the probability scores for matches (P), the cross correlation score (XC, another measure of the match quality), and the number of ions of the matched MS/MS spectra relative to the number of possible ions in the theoretical spectrum (Peptide (Hits), Ions column).

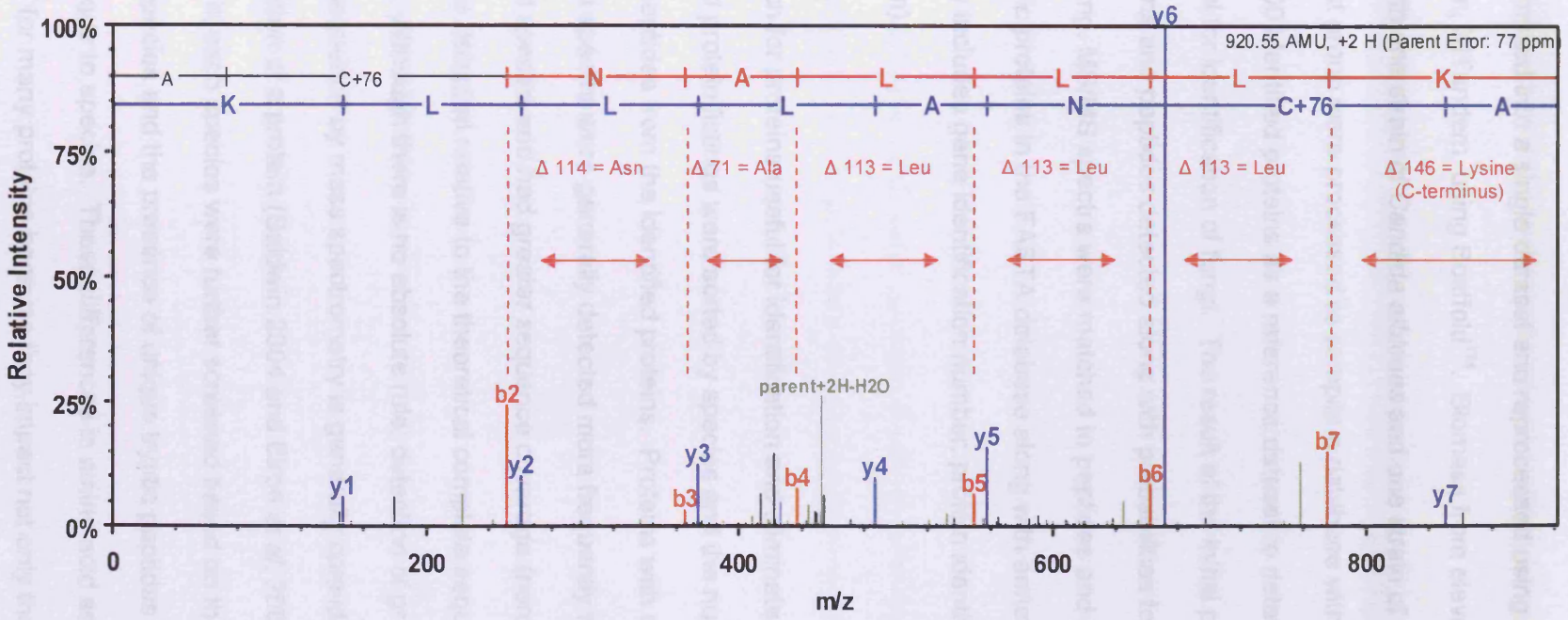


Figure 4.8. An example of how spectra are matched to a theoretical tryptic peptide from enolase (graphic generated by Scaffold™). Differences observed in the *m/z* of the ions of the MS/MS spectrum correspond to individual amino acids in the peptide chain. Ions of the y and b series, from breaks between the carbon and nitrogen of the peptide bond, predominate MS/MS spectra. Mass differences in the b series of ions, shown in red, correlate with the amino acid sequence '-N-A-L-L-L-K-'. A cysteine residue modified by hydroxyethyldisulfide (HED) in the protein extraction solution is shown in the theoretical sequence (both b and y series) as 'C+76'.

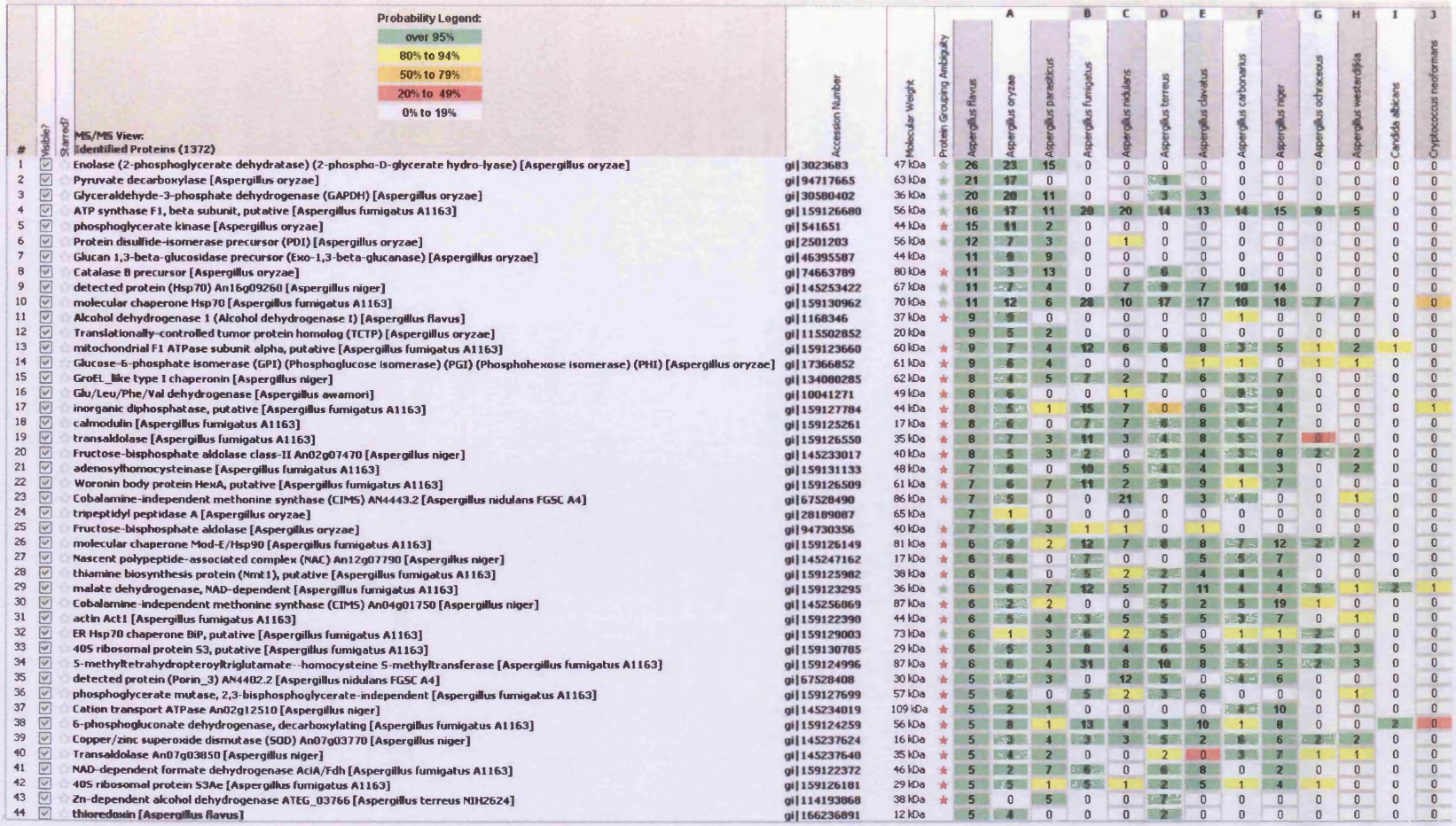
Results from 61 LC/MS/MS analyses of mycelia and spores from 29 reference strains were compiled into a single dataset and reprocessed using an additional matching algorithm, X!Tandem, using Scaffold™. Biomass from eleven *Aspergillus* species, along with one strain of *Candida albicans* and one strain of *Cryptococcus neoformans* as an out group were processed to compile a database with over 400,000 spectra and over 1300 identified proteins as a reference dataset to determine which proteins may be useful for identification of fungi. The result of the initial processing was a large list of proteins and peptides detected along with probabilities for correct matches. In the processing, MS/MS spectra were matched to peptides and those peptides were linked to specific proteins in the FASTA database along with annotation for those proteins (typically includes gene identification number, protein identification and source organism).

To search for proteins useful for identification and eliminate those that are not, the compiled protein listings were sorted by species and the number of spectra assigned to unique peptides from the identified proteins. Proteins with a greater number of assigned spectra were generally detected more frequently than those with fewer assigned spectra and had greater sequence coverage (percentage of amino acid sequence detected relative to the theoretical complete sequence) for the identified proteins. Although there is no absolute rule, detection of greater than 20% of the protein sequence by mass spectrometry is generally considered good coverage for identification of a protein (Baldwin 2004 and Biron *et al.* 2006). The relatively abundant proteins in each species were further screened based on the matching of the protein to correct species and the presence of unique tryptic peptides within the proteins that were unique to species. These differences in amino acid sequences were readily detected for many proteins because they impact not only the relative masses and

spectra of the peptides, but also cleavage sites within the proteins. As a general rule, tryptic peptides of proteins with sequences unique to species correctly identified fungi to the species or section level if four or more unique peptides were detected for a given protein and protein coverage was greater than 15%.

Identification by protein profiling is presented here largely from the perspective of *A. flavus* and *A. oryzae*, but a selection of proteins and peptides that may be useful for identifying Aspergilli in general are summarized with the *A. flavus/oryzae* examples presented below. An example view of the compiled dataset in Scaffold™, sorted on *Aspergillus flavus* and the number of unique peptides detected, is presented in Figure 4.9. In the example Scaffold™ table, note that some of the abundant proteins indentified for *A. flavus* are uniquely assigned to species in section Flavi (enolase, glyceraldehyde-3-phosphate dehydrogenase, pyruvate decarboxylase) and are candidate proteins for classifying Aspergilli. Other proteins, such as ATP synthase and molecular chaperone Hsp70, share enough sequence across species that they yield matches to multiple species within the genus and consequently cannot be used to classify Aspergilli to the species level.

Proteins such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH), enolase, ATP synthase, ribosomal proteins, and chaperonins dominated the profiles of the tryptic digests from Aspergilli. Other proteins, such as pyruvate decarboxylase, phosphoglycerate kinase, alcohol dehydrogenase, superoxide dismutase (Cu-Zn and Mn), catalyse, 14-3-3 family proteins, cobalamine-independent methionine synthase, translationally-controlled tumour protein (TCTP), triosephosphate isomerise (TPI), malate dehydrogenase, actin, tubulin, calmodulin, and Woronin body proteins along with various proteases, translation elongation and initiation factors, and polysaccharide degrading enzymes were frequently identified. Evaluations of specific detected proteins for classifying Aspergilli by protein profiling are summarized below.



Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

#	Visible?	Starred?	MS/MS View: Identified Proteins (1372)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	A Aspergillus flavus	A Aspergillus oryzae	A Aspergillus parasiticus	B Aspergillus fumigatus	C Aspergillus nidulans	D Aspergillus terreus	E Aspergillus clavatus	F Aspergillus carbonarius	F Aspergillus niger	G Aspergillus ochraceus	H Aspergillus westerdijkia	I Candida albicans	J Cryptococcus neoformans
1	☑	☐	Enolase (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase) [Aspergillus oryzae]	gi|3023683	47 kDa	☆	26	23	15	0	0	0	0	0	0	0	0	0	0
2	☑	☐	Pyruvate decarboxylase [Aspergillus oryzae]	gi|94717665	63 kDa	☆	21	17	0	0	0	1	0	0	0	0	0	0	0
3	☑	☐	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) [Aspergillus oryzae]	gi|30580402	36 kDa	☆	20	20	11	0	0	3	3	0	0	0	0	0	0
4	☑	☐	ATP synthase F1, beta subunit, putative [Aspergillus fumigatus A1163]	gi|159126680	56 kDa	☆	16	17	11	20	20	14	13	14	15	9	5	0	0
5	☑	☐	phosphoglycerate kinase [Aspergillus oryzae]	gi|541651	44 kDa	★	15	11	2	0	0	0	0	0	0	0	0	0	0
6	☑	☐	Protein disulfide-isomerase precursor (PDI) [Aspergillus oryzae]	gi|2501203	56 kDa	☆	12	7	3	0	1	0	0	0	0	0	0	0	0
7	☑	☐	Glucan 1,3-beta-glucosidase precursor (Exo-1,3-beta-glucanase) [Aspergillus oryzae]	gi|46395587	44 kDa		11	9	9	0	0	0	0	0	0	0	0	0	0
8	☑	☐	Catalase B precursor [Aspergillus oryzae]	gi|74663789	80 kDa	★	11	3	13	0	0	6	0	0	0	0	0	0	0
9	☑	☐	detected protein (Hsp70) An16g09260 [Aspergillus niger]	gi|145253422	67 kDa	☆	11	7	4	0	7	9	7	10	14	0	0	0	0
10	☑	☐	molecular chaperone Hsp70 [Aspergillus fumigatus A1163]	gi|159130962	70 kDa	☆	11	12	6	28	10	17	17	10	18	7	7	0	0
11	☑	☐	Alcohol dehydrogenase 1 (Alcohol dehydrogenase I) [Aspergillus flavus]	gi|1168346	37 kDa	★	9	5	0	0	0	0	0	1	0	0	0	0	0
12	☑	☐	Translationally-controlled tumor protein homolog (TCTP) [Aspergillus oryzae]	gi|115502852	20 kDa		9	5	2	0	0	0	0	0	0	0	0	0	0
13	☑	☐	mitochondrial F1 ATPase subunit alpha, putative [Aspergillus fumigatus A1163]	gi|159123660	60 kDa	★	9	7	4	12	6	6	8	3	5	1	2	1	0
14	☑	☐	Glucose-6-phosphate isomerase (GPI) (Phosphoglucose isomerase) (PGI) (Phosphohexose isomerase) (PHI) [Aspergillus oryzae]	gi|17366852	61 kDa	★	9	6	4	0	0	0	1	1	0	1	1	0	0
15	☑	☐	GroEL_like type I chaperonin [Aspergillus niger]	gi|134080285	62 kDa	★	8	4	5	7	2	7	6	3	7	0	0	0	0
16	☑	☐	Glu/Leu/Phe/Val dehydrogenase [Aspergillus awamori]	gi|10041271	49 kDa	★	8	6	0	0	1	0	0	0	9	0	0	0	0
17	☑	☐	inorganic diphosphatase, putative [Aspergillus fumigatus A1163]	gi|159127784	44 kDa	★	8	5	1	15	7	0	6	3	4	0	0	0	1
18	☑	☐	calmodulin [Aspergillus fumigatus A1163]	gi|159125261	17 kDa	★	8	6	0	7	7	6	8	6	7	0	0	0	0
19	☑	☐	transaldolase [Aspergillus fumigatus A1163]	gi|159126550	35 kDa	★	8	7	3	11	3	4	8	5	7	0	0	0	0
20	☑	☐	Fructose-bisphosphate aldolase class-II An02g07470 [Aspergillus niger]	gi|145233017	40 kDa	★	8	5	3	2	0	5	4	3	8	2	2	0	0
21	☑	☐	adenosylhomocysteinase [Aspergillus fumigatus A1163]	gi|159131133	48 kDa	★	7	6	0	10	5	4	4	4	3	0	2	0	0
22	☑	☐	Woronin body protein HexA, putative [Aspergillus fumigatus A1163]	gi|159126509	61 kDa	★	7	6	7	11	2	9	9	1	7	0	0	0	0
23	☑	☐	Cobalamine-independent methonine synthase (CIMS) AN4443.2 [Aspergillus nidulans FGSC A4]	gi|67528490	86 kDa	★	7	5	0	0	21	0	3	4	0	0	1	0	0
24	☑	☐	tripeptidyl peptidase A [Aspergillus oryzae]	gi|28189087	65 kDa		7	1	0	0	0	0	0	0	0	0	0	0	0
25	☑	☐	Fructose-bisphosphate aldolase [Aspergillus oryzae]	gi|94730356	40 kDa	★	7	6	3	1	1	0	1	0	0	0	0	0	0
26	☑	☐	molecular chaperone Mod-E/Hsp90 [Aspergillus fumigatus A1163]	gi|159126149	81 kDa	★	6	9	2	12	7	8	8	7	12	2	2	0	0
27	☑	☐	Nascent polypeptide-associated complex (NAC) An12g07790 [Aspergillus niger]	gi|145247162	17 kDa	★	6	6	0	7	0	0	5	5	7	0	0	0	0
28	☑	☐	thiamine biosynthesis protein (Nmt1), putative [Aspergillus fumigatus A1163]	gi|159125982	38 kDa	★	6	4	0	5	2	2	4	4	4	0	0	0	0
29	☑	☐	malate dehydrogenase, NAD-dependent [Aspergillus fumigatus A1163]	gi|159123295	36 kDa	☆	6	6	7	12	5	7	11	4	4	5	1	2	1
30	☑	☐	Cobalamine-independent methonine synthase (CIMS) An04g01750 [Aspergillus niger]	gi|145256869	87 kDa	★	6	2	2	0	0	5	2	5	19	1	0	0	0
31	☑	☐	actin Act1 [Aspergillus fumigatus A1163]	gi|159122390	44 kDa	★	6	5	4	3	5	5	5	3	7	0	1	0	0
32	☑	☐	ER Hsp70 chaperone BiP, putative [Aspergillus fumigatus A1163]	gi|159129003	73 kDa	☆	6	1	3	6	2	5	0	1	1	2	0	0	0
33	☑	☐	40S ribosomal protein S3, putative [Aspergillus fumigatus A1163]	gi|159130785	29 kDa	★	6	5	3	8	4	6	5	4	3	2	3	0	0
34	☑	☐	5-methyltetrahydropteroyltriglutamate--homocysteine 5-methyltransferase [Aspergillus fumigatus A1163]	gi|159124996	87 kDa	★	6	6	4	31	8	10	8	5	5	2	3	0	0
35	☑	☐	detected protein (Porin_3) AN4402.2 [Aspergillus nidulans FGSC A4]	gi|67528408	30 kDa	★	5	2	3	0	12	5	0	4	6	0	0	0	0
36	☑	☐	phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent [Aspergillus fumigatus A1163]	gi|159127699	57 kDa	★	5	6	0	5	2	3	6	0	0	0	1	0	0
37	☑	☐	Cation transport ATPase An02g12510 [Aspergillus niger]	gi|145234019	109 kDa	★	5	2	1	0	0	0	0	4	10	0	0	0	0
38	☑	☐	6-phosphogluconate dehydrogenase, decarboxylating [Aspergillus fumigatus A1163]	gi|159124259	56 kDa	★	5	8	1	13	4	3	10	1	8	0	0	2	0
39	☑	☐	Copper/zinc superoxide dismutase (SOD) An07g03770 [Aspergillus niger]	gi|145237624	16 kDa	★	5	3	4	3	3	5	2	6	6	2	2	0	0
40	☑	☐	Transaldolase An07g03850 [Aspergillus niger]	gi|145237640	35 kDa	★	5	4	2	0	0	2	0	3	7	1	1	0	0
41	☑	☐	NAD-dependent formate dehydrogenase AciA/Fdh [Aspergillus fumigatus A1163]	gi|159122372	46 kDa	★	5	2	7	6	0	6	8	0	2	0	0	0	0
42	☑	☐	40S ribosomal protein S3Ae [Aspergillus fumigatus A1163]	gi|159126181	29 kDa	★	5	5	1	5	1	2	5	1	4	1	0	0	0
43	☑	☐	Zn-dependent alcohol dehydrogenase ATEG_03766 [Aspergillus terreus NIH2624]	gi|114193868	38 kDa	★	5	0	5	0	0	7	0	0	0	0	0	0	0
44	☑	☐	thioredoxin [Aspergillus flavus]	gi|166236891	12 kDa		5	4	0	0	0	2	0	0	0	0	0	0	0

Figure 4.9. An example view of a Scaffold™ results table compiled from LC/MS/MS analyses of different species of *Aspergillus*, as well as *C. albicans* and *C. neoformans*. Results have been sorted on *A. flavus* and the number of unique peptides detected. The numbers represent the number of spectra matched to unique peptides in the protein and those in green indicate high probability of correct identification ($p<0.05$).

ATP synthase. ATP synthase was a frequently detected protein and protein coverage was typically very good with coverage for protein identified as ATP synthase F1, beta subunit, ranging from approximately 10-50%. However, ATP synthase did not match samples to the correct species. In the example Scaffold™ table, spectra for a number of species matched to the listing for ATP synthase from A. fumigatus (Figures 4.9, row 4 and 4.10). Although frequently detected with good sequence coverage, tryptic peptides from ATP synthase were not useful for classification because there is considerable sequence similarity across *Aspergillus* species (>90% identity between *A. fumigatus*, *A. flavus*, *A. oryzae*, *A. clavatus*, *A. terreus*, and *A. clavatus*).

Figure 4.10. An example view (Scaffold™) of sequence coverage for spectra from individual samples matched to peptides of ATP synthase F1, beta subunit, from *Aspergillus fumigatus*. The first column shows the regions of matched peptides, in yellow, superimposed over the total theoretical amino acid sequence. Areas that did not have matching spectra are white. Sequence layout, left to right, is from N-terminus to C-terminus. Column two lists the individual samples that yielded the peptides and column three lists the number of peptides in the sample that matched peptides in *A. fumigatus*. Tryptic digests of all *Aspergillus* species tested yielded matches to *A. fumigatus* ATP synthase and few unique sequences from ATP synthase were detected that would uniquely identify the organisms to the correct species. Samples are from mycelia; samples marked with '*' are from conidia.

Sequence Coverage	Sample	#Pep	%Cov
	Aspergillus carbonarius 369	8	30%
	Aspergillus carbonarius 369*	11	35%
	Aspergillus clavatus 3538	9	28%
	Aspergillus clavatus 1	8	24%
	Aspergillus clavatus 1*	7	21%
	Aspergillus clavatus 1	10	28%
	Aspergillus clavatus 3538	9	25%
	Aspergillus clavatus 3538*	10	28%
	Aspergillus flavus 3357	8	24%
	Aspergillus flavus 1957	9	33%
	Aspergillus flavus 3357	10	28%
	Aspergillus flavus 485	12	42%
	Aspergillus flavus 3646*	7	26%
	Aspergillus flavus 827*	9	30%
	Aspergillus flavus 1957	9	25%
	Aspergillus fumigatus Af293	18	53%
	Aspergillus fumigatus 163	14	42%
	Aspergillus fumigatus 163*	9	29%
	Aspergillus fumigatus Af293*	11	31%
	Aspergillus fumigatus Af293	14	42%
	Aspergillus nidulans A4*	14	39%
	Aspergillus nidulans 187	8	23%
	Aspergillus nidulans 187	14	38%
	Aspergillus nidulans 187*	15	47%
	Aspergillus nidulans A4	10	32%
	Aspergillus nidulans 326	11	33%
	Aspergillus nidulans 326	12	35%
	Aspergillus nidulans 326*	9	32%
	Aspergillus ochraceous 398*	9	23%
	Aspergillus oryzae 5590*	11	36%
	Aspergillus oryzae 1988*	6	20%
	Aspergillus oryzae RIB40	6	19%
	Aspergillus oryzae 4823*	7	24%
	Aspergillus oryzae 5590	9	30%
	Aspergillus oryzae RIB128*	6	24%
	Aspergillus oryzae 447	10	27%
	Aspergillus oryzae 5590	3	8.1%
	Aspergillus parasiticus 502*	9	29%
	Aspergillus parasiticus SU1	8	24%
	Aspergillus parasiticus 502	5	15%
	Aspergillus parasiticus 502	3	9.8%
	Aspergillus terreus 274*	3	11%
	Aspergillus terreus 274	4	12%
	Aspergillus terreus 255*	2	6.7%
	Aspergillus terreus 1913	11	34%
	Aspergillus terreus 255	3	7.3%
	Aspergillus terreus 274	4	12%
	Aspergillus terreus 1913*	4	16%
	Aspergillus terreus 255	8	22%
	Aspergillus terreus 1913	11	34%
	Aspergillus westerdijkiae 5175*	4	11%
	Aspergillus westerdijkiae 3134*	3	7.9%
	Clayton 270	8	23%

Tubulin and actin. Tubulin and actin were not major proteins detected in the extracts, often with 3 peptides or fewer detected in samples, and the sequence coverages obtained were often less than 20%. β-Tubulin was detected only in three *A. clavatus* samples, but peptides from tubulin alpha-1 subunit were detected in several samples. An example of the peptide coverage for tubulin alpha-1 subunit for *A. fumigatus* and the shared sequences of the *A. fumigatus* peptides with other Aspergilli are presented in Figure 4.11. Note that a tryptic peptide from tubulin alpha-1 subunit of *A. fumigatus* with amino acid sequence 'TIYCDLEPNVVDEV**R**' is shared with *A. clavatus* and *A. terreus* and may also share peptides 'YMATCLLY**R**' and 'TIQFVDWCPTGF**K**' with *A. clavatus.* Similarly, relatively few peptides from actin were detected by profiling and some of those were shared across species, such as 'YPIEHGVVTNWDDME**K**', and 'SYELPDGQVITIGNE**R**' as indicated in Figure 4.12.

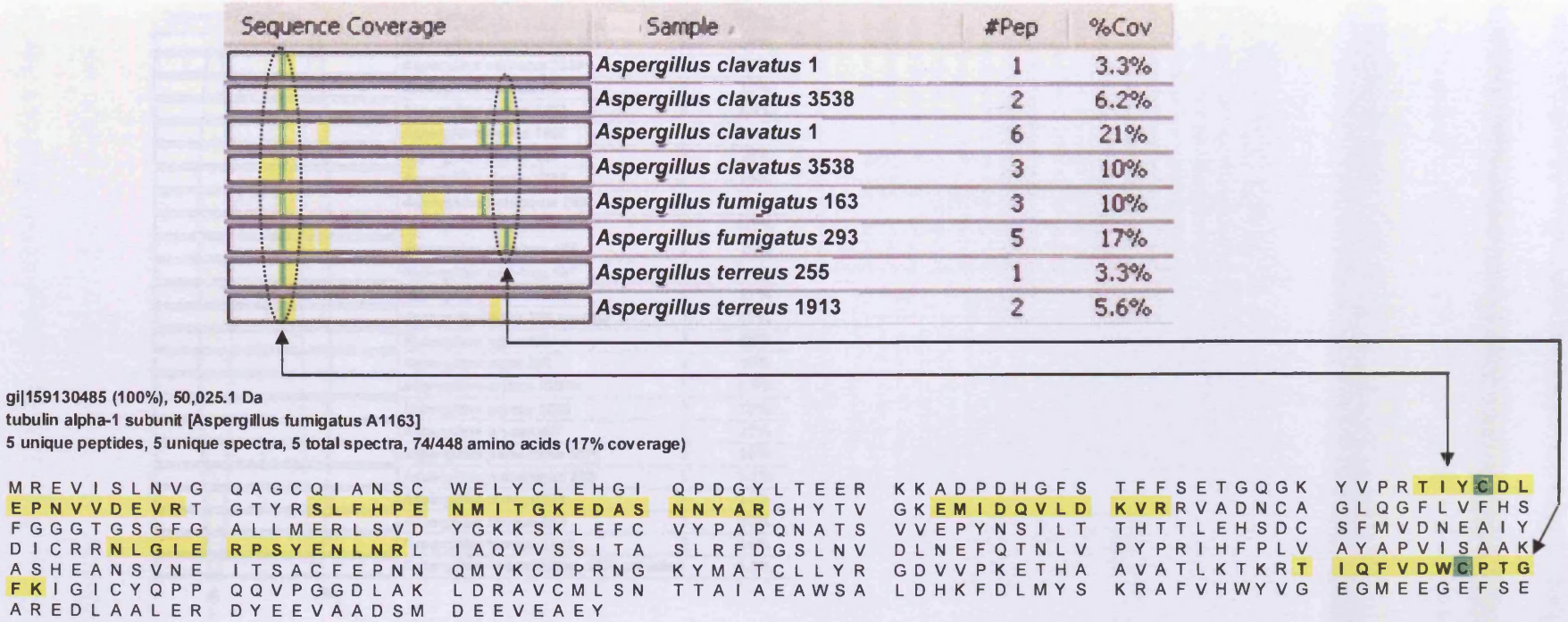


Figure 4.11. An example of the peptide coverage for tubulin alpha-1 subunit for *A. fumigatus* and the shared sequences across other *Aspergillus* species. Relatively few peptides from tubulin were detected and a number of those were shared across *A. fumigatus*, *A. terreus* and *A. clavatus* (top). Peptides detected in *A. fumigatus* sample, shown in yellow, superimposed on the amino acid sequence of tubulin; modified cysteine residues shown in green (bottom). Shared sequences indicated on the sequence coverage plot are shown matched to the corresponding amino acid sequence of tubulin alpha-1 subunit from *A. fumigatus*.

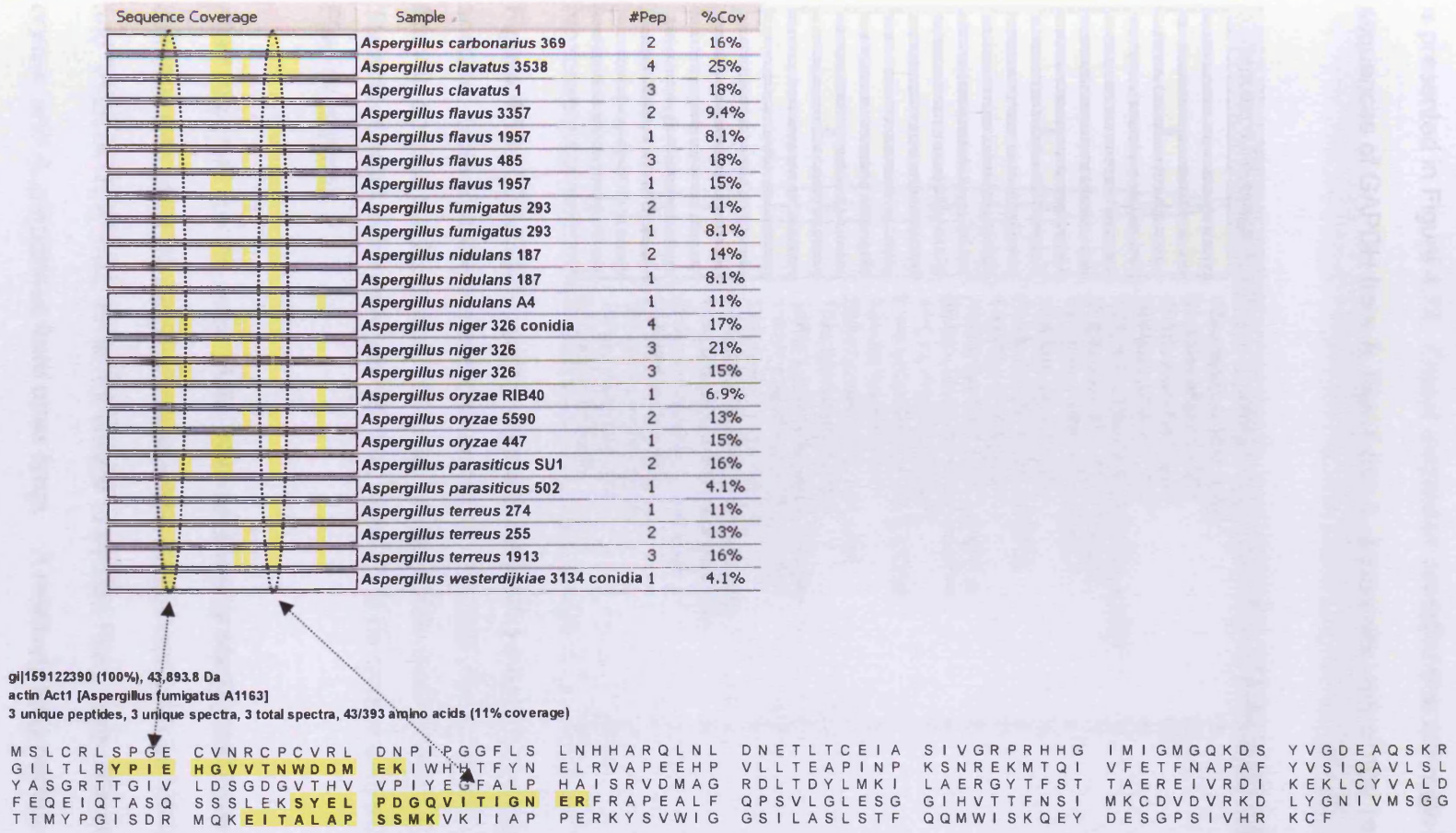


Figure 4.12. An example of the peptide coverage for matches across *Aspergillus* species to actin from *A. fumigatus*. Relatively few peptides from actin were detected and a number of those were shared across many of the *Aspergillus* species (top). Shared sequences indicated on the sequence coverage plot are shown matched to the corresponding amino acid sequence of actin from *A. fumigatus*. (bottom).

amino acid sequences within GAPDH, but the amino acid sequence is specific to section Flavi (Table 4.2). The alignment of complete GAPDH sequences from the test species and the region with tryptic peptides diagnostic for species is presented in Figure 4.14.

Genus species	tryptic peptides
Aspergillus clavatus	GILGYTEDAIVSSDVNGDE**R**SSVFDAAAGISLNPNFV**K**
Aspergillus flavus/oryzae	GILGYTEDDIVSTDLIGDA**H**SSIFDA**K**AGIALNEHFI**K**
Aspergillus fumigatus	NILGYTEDDVVSSDLNGDE**R**SSIFDA**K**AGISLNPNFV**K**
Aspergillus nidulans	GILGYTEDDIVSTDLNGDT**R**SSIFDA**K**AGIALNSNFI**K**
Aspergillus niger	GILGYTEDDIVSSDLNGDDHSSIFDA**K**AGIALNSNFV**K**
Aspergillus terreus	GILGYTEDEVVSTDLNGDD**R**SSIFDA**K**AGIALNEHFV**K**

Table 4.2. Region within GAPDH that spans tryptic peptides diagnostic for identification of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**. Note that in this region the sequence for *A. flavus/oryzae* differs from other species by a few amino acids and an arginine cleavage site (highlighted in red).

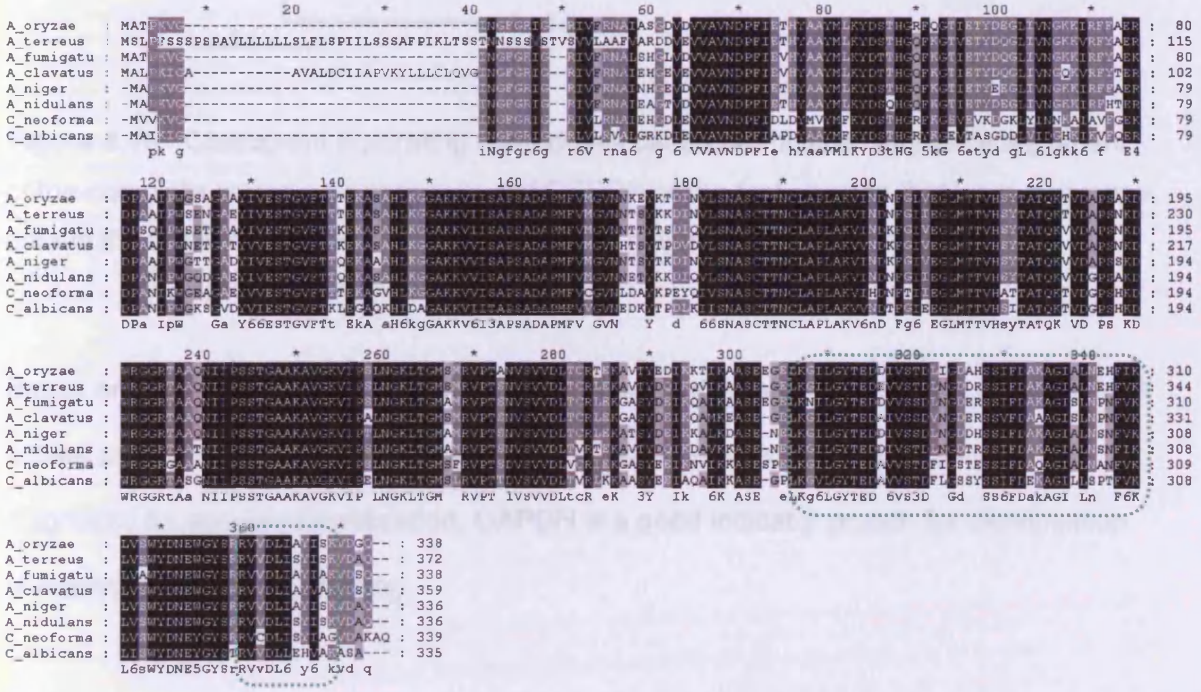


Figure 4.14. Multiple alignment of the complete amino acid sequences from GAPDH (NCBI database) showing conserved and variable regions between test species. Regions with tryptic peptides diagnostic for classifying Aspergilli to species (Table 4.2) are outlined in green.

Although complete sequence coverage is not achieved by profiling, the amino acid sequence of GAPDH has a number of variable regions that are differential for species and may be a suitable genetic character for classifying fungi as illustrated in Figure 4.15.

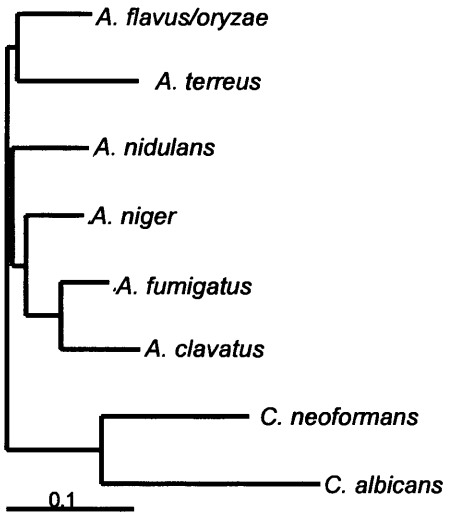


Figure 4.15. Cladogram illustrating relatedness based on multiple sequence alignment of the complete amino acid sequences of GAPDH of the test species that have sequence data in the NCBI database (tree prepared using neighbour joining method).

Based on the relative abundance of GAPDH, the good sequence coverage obtained for it in total protein digests and the frequent detection of peptides from GAPDH that are diagnostic for species identification, GAPDH is a good indicator protein for identification of Aspergilli by protein profiling.

Enolase. Enolase (47 kDa) was an abundant protein with good sequence coverage (Figure 14.16) and was generally good for classifying the test samples correctly to section or species, but a protein identified as enolase/allergen Asp F 22 from *A. fumigatus* sometimes matched peptides across several other species, particularly

A. clavatus. Consequently, classification of Aspergilli by identification of peptides from enolase should only be used as supporting evidence unless a unique peptide such as 'IEEELGNNAIYAGE**K**' is detected with high confidence, Table 4.3. There are other tryptic peptides that may differentiate species, but many of the variable regions within enolase are rich in lysine which results in very short peptides, less valuable for identification by MS/MS. The alignment of complete enolase amino acid sequences from the test species and the region with tryptic peptides diagnostic for species is presented in Figure 4.17.

Sequence Coverage	Sample	# Peptides	%Coverage
	01AflavusNRRL1957	19	67%
	09AflavusNRRL3357	16	60%
	09AflavusNRRL3357	7	24%
	22AflavusNRRL3646 (conidia)	6	21%
	15AflavusNRRL1957	16	58%
	20AflavusPapa827 (conidia)	8	29%
	10AoryzaeNRRL447	10	41%
	05AoryzaeRIB128 (conidia)	6	20%
	06AoryzaeNRRL5590	3	9.4%
	13AoryzaeNRRL5590	17	66%
	13AoryzaeNRRL1988 (conidia)	5	18%
	11AoryzaeNRRL5590 (conidia)	12	41%
	01AoryzaeNRRL4823 (conidia)	6	21%
	A_parasiticus_myc01_NRRL502	7	31%
	17AparasiticusNRRL424	6	22%
	12AparasiticusNRRL502	6	20%
	13AparasiticusSU1	11	39%
	18AparasiticusNRRL502 (conidia)	7	27%
	01Clayton270myc	23	67%

Figure 4.16. LC/MS/MS peptide coverage for enolase from *A. oryzae* across test strains. In this example, only species from section Flavi (*A. flavus*, *A. oryzae* and *A. parasiticus*) are matched to enolase from *A. oryzae*. Coverage is generally very good with sequence coverage as high as 67%.

Genus species	tryptic peptides
Aspergillus clavatus	TGAPA**R**SER**L**A**K**LNQIL**R**IEEELGDNAVYAGD**K**F**R**TAVNL
Aspergillus flavus/oryzae	TGAPA**R**SER**L**A**K**LNQIL**R**IEEELGNNAIYAGE**K**F**R**TSVNL
Aspergillus fumigatus	TGAPC**R**SER**L**A**K**LNQIL**R**IEEELGENAVYAGS**K**F**R**TAVNL
Aspergillus nidulans	TGAPA**R**SER**L**A**K**LNQIL**R**IEEELGENAVYAGQNF**RK**SVNL
Aspergillus niger	TGAPA**R**SER**L**A**K**LNQIL**R**IEEELGDNAVYAGE**K**F**R**TAVNL
Aspergillus terreus	TGAPA**R**SER**L**A**K**LNQIL**R**IEEELGDNAVYAGE**K**F**R**TAVNL

Table 4.3. Region within enolase that spans tryptic peptides diagnostic for identification of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**.

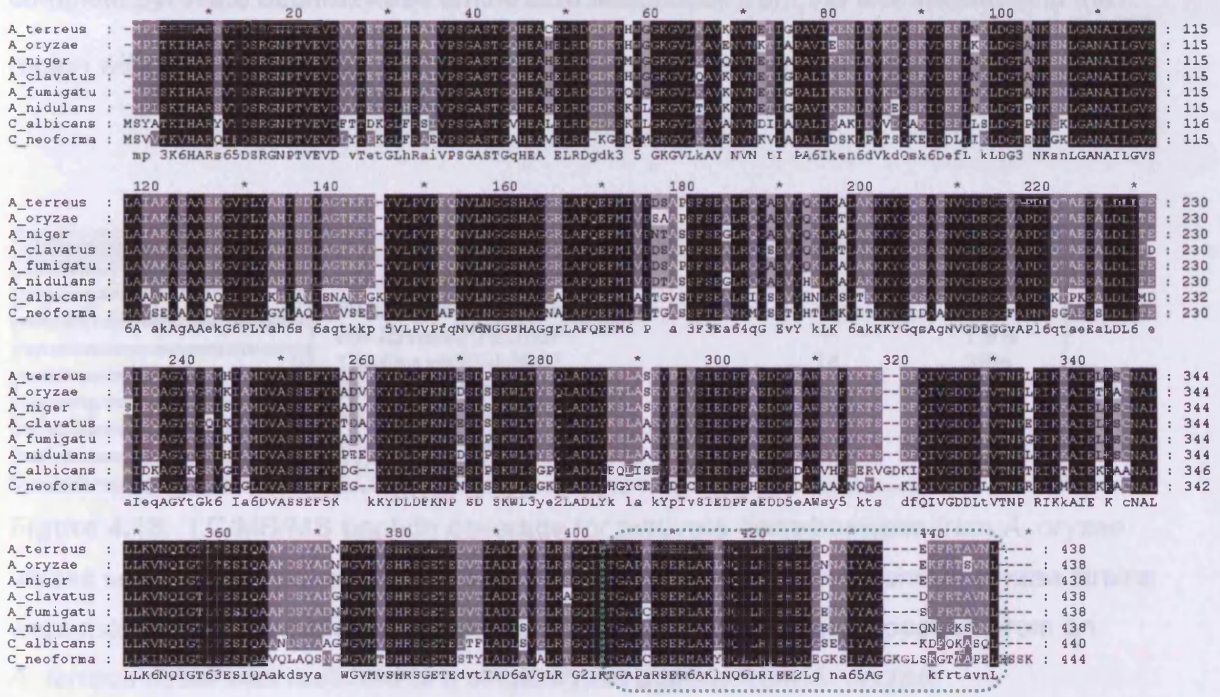


Figure 4.17. Multiple alignment of complete enolase amino acid sequences (NCBI) from test species showing variable regions between species. A region with tryptic peptides diagnostic for classifying Aspergilli to species (Table 4.3) is outlined in green.

Pyruvate decarboxylase. Pyruvate decarboxylase (63 kDa) was not an abundant protein, but in those samples with reasonable coverage (Figure 4.18) there were several variable regions within tryptic peptides that were useful for classifying organisms to the correct section or species. Peptides 'YIHGWEAVYNDIQPWDFLNIPVAFGA**K**' (Table 4.4), 'ELFANEEFASAPCLQLVELHMP**R**', 'MGNLNVGPVSPPSNLLPDNE**K**', and

'PVYISLPTDMVT**K**' from *A. flavus* and *A. oryzae* as well as peptides from the corresponding regions of other fungi tested are very diagnostic for classification by LC/MS/MS. Tryptic peptides of pyruvate decarboxylase performed well for classification across *Aspergillus* species and were suitable for identification *C. albicans* as well. Based on these results, pyruvate decarboxylase is a good candidate for follow-up targeted protein analysis and DNA sequencing for other species of interest that do not currently have gene or EST sequence available. The multiple sequence alignment of complete pyruvate decaroxylase amino acid sequences from the test species and the region with tryptic peptides diagnostic for species is presented in Figure 4.19.

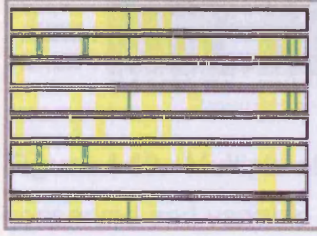
Sequence Coverage	Sample	# Peptides	%Coverage
	01AflavusNRRL1957	10	26%
	09AflavusNRRL3357	16	48%
	09AflavusNRRL3357	1	1.9%
	15AflavusNRRL1957	14	38%
	04AflavusNRRL485	7	17%
	13AoryzaeNRRL5590	17	54%
	14AterreusNRRL274	1	4.7%
	01Clayton270myc	12	33%

Figure 4.18. LC/MS/MS peptide coverage for pyruvate decarboxylase from *A. oryzae* across several test strains. In this example, spectra from *A. flavus* and *A. oryzae* strains were matched to pyruvate decarboxylase from *A. oryzae*, but one spectrum from an *A. terreus* strain was matched to a single tryptic peptide from *A. oryzae*.

Genus species	tryptic peptides
Aspergillus clavatus	YIHGWDAGYNDIQEWDN**K**NIPTVFGGGDFY**K**
Aspergillus flavus/oryzae	YIHGWEAVYNDIQPWDFLNIPVAFGA**K**
Aspergillus fumigatus	YIHGWEASYNDIQQWDY**K**SLPVAFGAG**K**
Aspergillus nidulans	FIHGWDESYNDIQTWDI**K**GLPVAFGG**K**
Aspergillus niger	YIHGWDESYNDIQPWDIEGLP**R**VFGA**K**
Aspergillus terreus	FIHGWDEAYNDIQPWDI**K**GLPVVFGA**K**

Table 4.4. Region within pyruvate decarboxylase that spans tryptic peptides diagnostic for identification of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**. In this region, pyruvate decarboxylate differs from other species at several amino acids and one lysine cleavage site (marked in red).

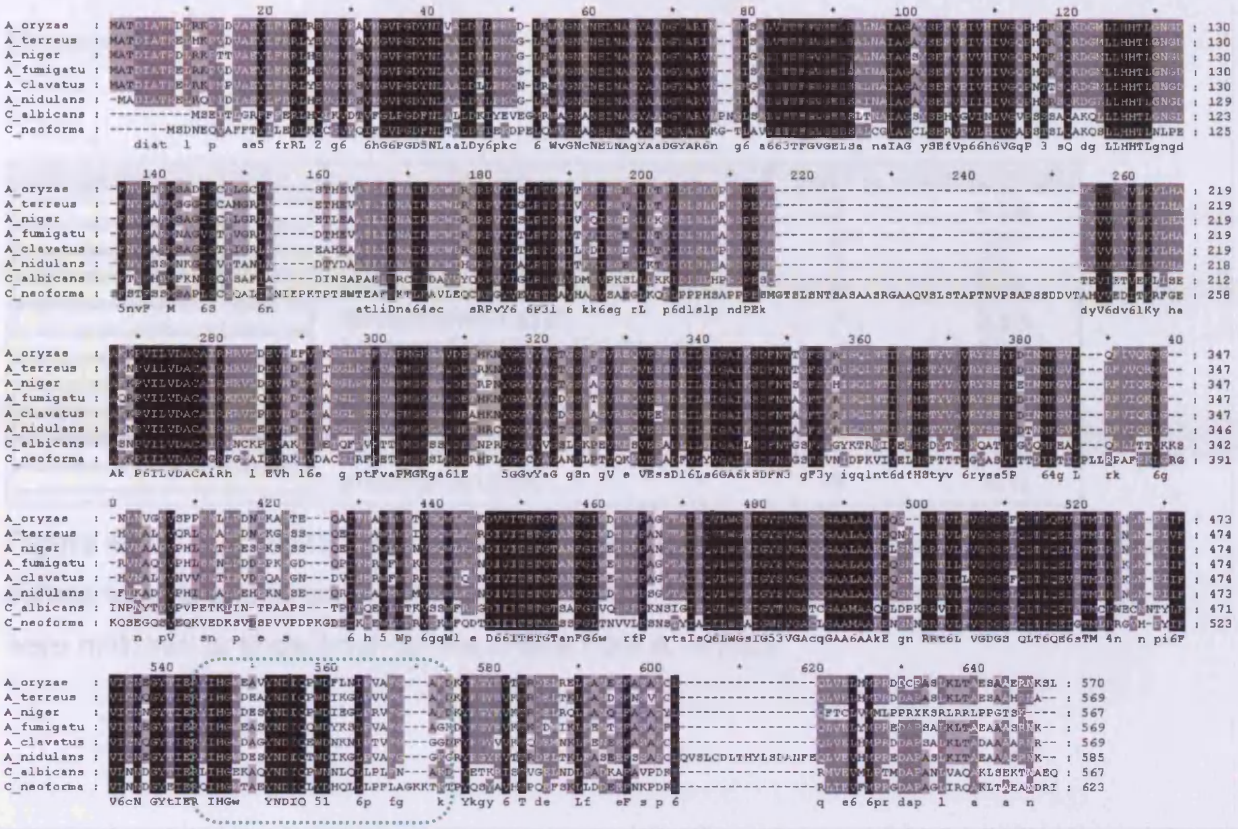


Figure 4.19. Multiple alignment of pyruvate decarboxylase sequences showing conserved and variable regions between test species. A region with tryptic peptides diagnostic for classifying Aspergilli to species (Table 4.4) is outlined in green.

Phosphoglycerate kinase. Phosphoglycerate kinase (44 kDa), though not as widely detected as pyruvate decarboxylase, yielded a number of peptides that could be used to classify Aspergilli. Sequence coverage across test samples for phosphoglycerate kinase from *A. oryzae* is presented in Figure 14.20. There are several regions of variability within pyruvate decarboxylate that are differential for species or section, but a number of these are lysine rich and yield small peptides with 6 or fewer amino acid residues. Tryptic peptides 'TILWNGPPGVFELEPFANAT**K**' and 'TGYATDADGIPDGYMGLDVGE**K**' from section Flavi and the peptides from corresponding regions of other Aspergilli are often detected and are diagnostic for classification of the fungi (Table 4.5). A multiple sequence alignment of complete phosphoglycerate kinase amino acid sequences from

the test species and the region with tryptic peptides diagnostic for species is presented in Figure 4.21.

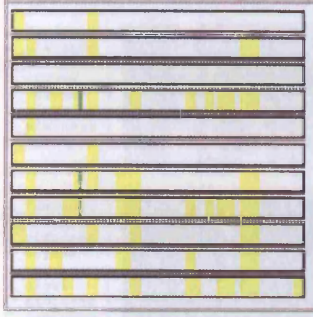
Sequence Coverage	Sample	# Peptides	%Coverage
	20AflavusPapa827 (conidia)	2	5.3%
	11AoryzaeNRRL5590 (conidia)	5	13%
	18AparasiticusNRRL502 (conidia)	0	0%
	13AoryzaeNRRL5590	9	31%
	09AflavusNRRL3357	1	2.9%
	17AparasiticusNRRL424 (conidia)	2	5.3%
	15AflavusNRRL1957	7	22%
	09AflavusNRRL3357	9	32%
	01AflavusNRRL1957	4	14%
	04AflavusNRRL485	5	18%
	01Clayton270myc	9	30%

Figure 4.20. LC/MS/MS peptide coverage for phosphoglycerate kinase from *A. oryzae* across several test strains. In this example, only spectra from species in section Flavi were matched to phosphoglycerate kinase from *A. oryzae*.

Genus species	tryptic peptides
Aspergillus clavatus	VGYATDEQGIPDGLMGLDVGQ**K**SVALY**K**ETIAEA**K**TILWNGPPGVFELEPFANGT**K**
Aspergillus flavus/oryzae	TGYATDADGIPDGYMGLDVGE**K**SVELY**KK**TIAEA**K**TILWNGPPGVFELEPFANAT**K**
Aspergillus fumigatus	TGYATDEEGIPDGYMGLDVGD**K**SV**K**LY**K**ETIAEA**K**TILWNGPPGVFEMEPFANGT**K**
Aspergillus nidulans	TGYATDEQGIPDGYMGLDVGE**K**SVESY**K**QTIAES**K**TILWNGPPGVFEMEPFA**K**AT**K**
Aspergillus niger	TGTATDAEGIPDGYMGLDVGE**K**SVELY**K**QTIAEA**K**TILWNGPPGVFELEPFANGT**K**
Aspergillus terreus	TGYATDADGIPDGFMGLDVGE**K**SVELY**K**QTIAEA**K**TILWNGPCGVFEMEPFANGT**K**

Table 4.5. Region within phosphoglycerate kinase with tryptic peptides diagnostic for identification of Aspergilli. Cleavage sites at lysine residues are in **BOLD**.

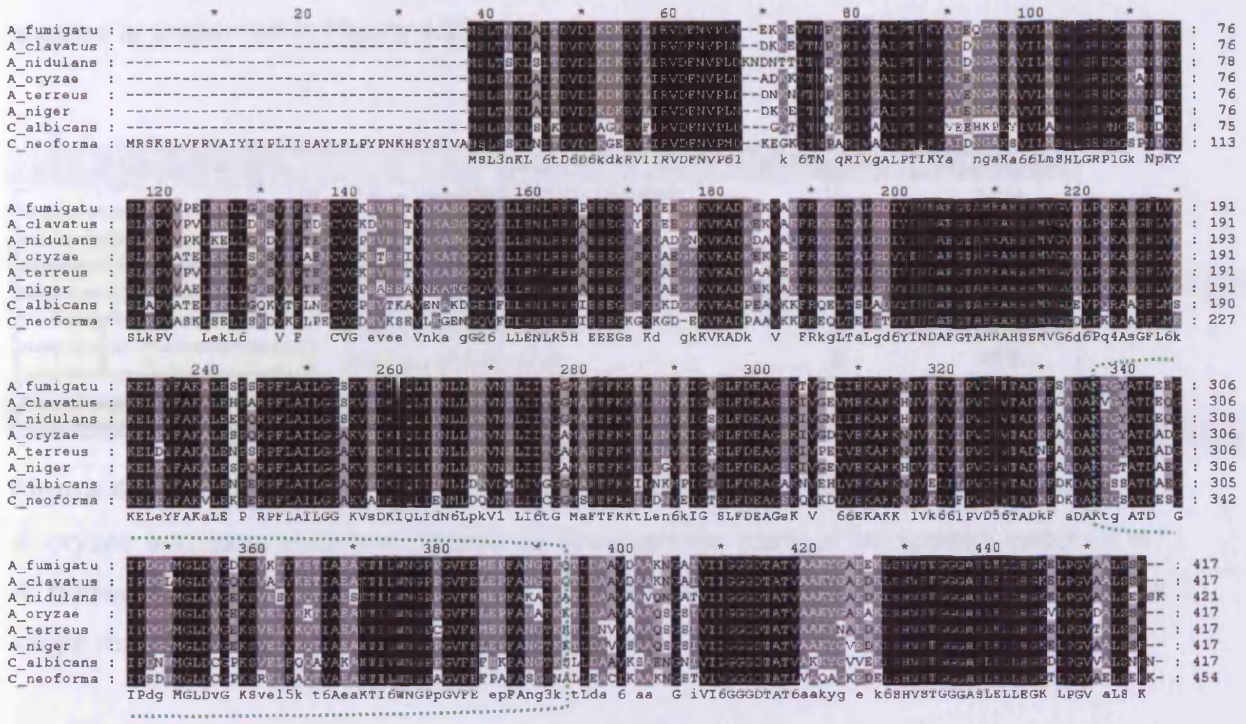
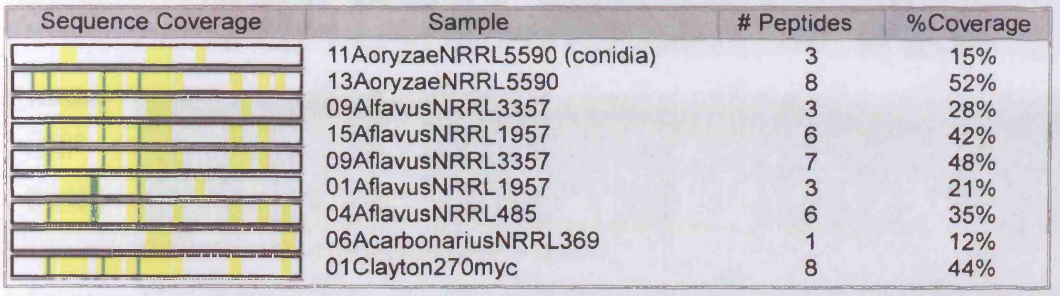


Figure 4.21. Multiple sequence alignment of phospoglycerate kinase sequences showing conserved and variable regions between test species. A region spanning tryptic peptides diagnostic for classifying Aspergilli to species (Table 4.5) is outlined in green.

Alcohol dehydrogenase. Zinc-binding alcohol dehydrogenase (37 kDa) yielded multiple tryptic peptides that were readily detected by LC/MS/MS profiling and matched with high confidence to alcohol dehydrogenase. Sequence coverage across test samples for zinc-binding alcohol dehydrogenase from *A. oryzae* is presented in Figure 14.22. Few were highly diagnostic for speciation, but one variable region yielded a relatively long tryptic peptide that if detected was specific to section or species - 'QADEPLCPNASLSGYTVDGTFQQYAIG**K**', Table 4.6. On review by multiple sequence alignment, alternate variable regions diagnostic for species would be generated by digesting alcohol dehydrogenase with 2% formic acid in water (cleaves at the carboxylic acid side of aspartic acid residues). This has not yet been confirmed experimentally. The multiple sequence alignment of complete alcohol dehydrogenase amino acid

sequences from the test species and the region with tryptic peptides diagnostic for species is presented in Figure 4.23.



Sequence Coverage	Sample	# Peptides	%Coverage
	11AoryzaeNRRL5590 (conidia)	3	15%
	13AoryzaeNRRL5590	8	52%
	09AlfavusNRRL3357	3	28%
	15AflavusNRRL1957	6	42%
	09AflavusNRRL3357	7	48%
	01AflavusNRRL1957	3	21%
	04AflavusNRRL485	6	35%
	06AcarbonariusNRRL369	1	12%
	01Clayton270myc	8	44%

Figure 4.22. LC/MS/MS peptide coverage for zinc-binding alcohol dehydrogenase from *A. oryzae* across several test strains. In this example, most of the spectra matched to *A. oryzae* alcohol dehydrogenase were from samples of *A. flavus* and *A. oryzae*, with a single match to a spectrum from an *A. carbonarius* sample.

Genus species	tryptic peptides
Aspergillus clavatus	WLNGSCLACEFC**K**QAEDPLCPHALLSGYTVDGTFQQYAIA**K**
Aspergillus flavus/oryzae	WLNGSCLACEFC**K**QADEPLCPNASLSGYTVDGTFQQYAIG**K**
Aspergillus fumigatus	WLNGSCLACEFC**K**QADEPLCQNALLSGYTVDGTFQQYTIG**K**
Aspergillus nidulans	WLNGSCGECEFC**R**QSDDPLCA**R**AQLSGYTVDGTFQQYALG**K**
Aspergillus niger	WLNGSCLACEFC**K**QAEEPLCPHALLSGYTVDGTFQQYAIA**K**
Aspergillus terreus	WLNGSCLSCEFCQQAQEPLCPNAQLSGYTVDGTFQQYAIG**K**

Table 4.6. Region within Zn-binding alcohol dehydrogenase with tryptic peptides diagnostic for speciation of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**. Note that *A. nidulans* and *A. terreus* have amino acid substitutions and altered cleavage sites in this region (marked in red).

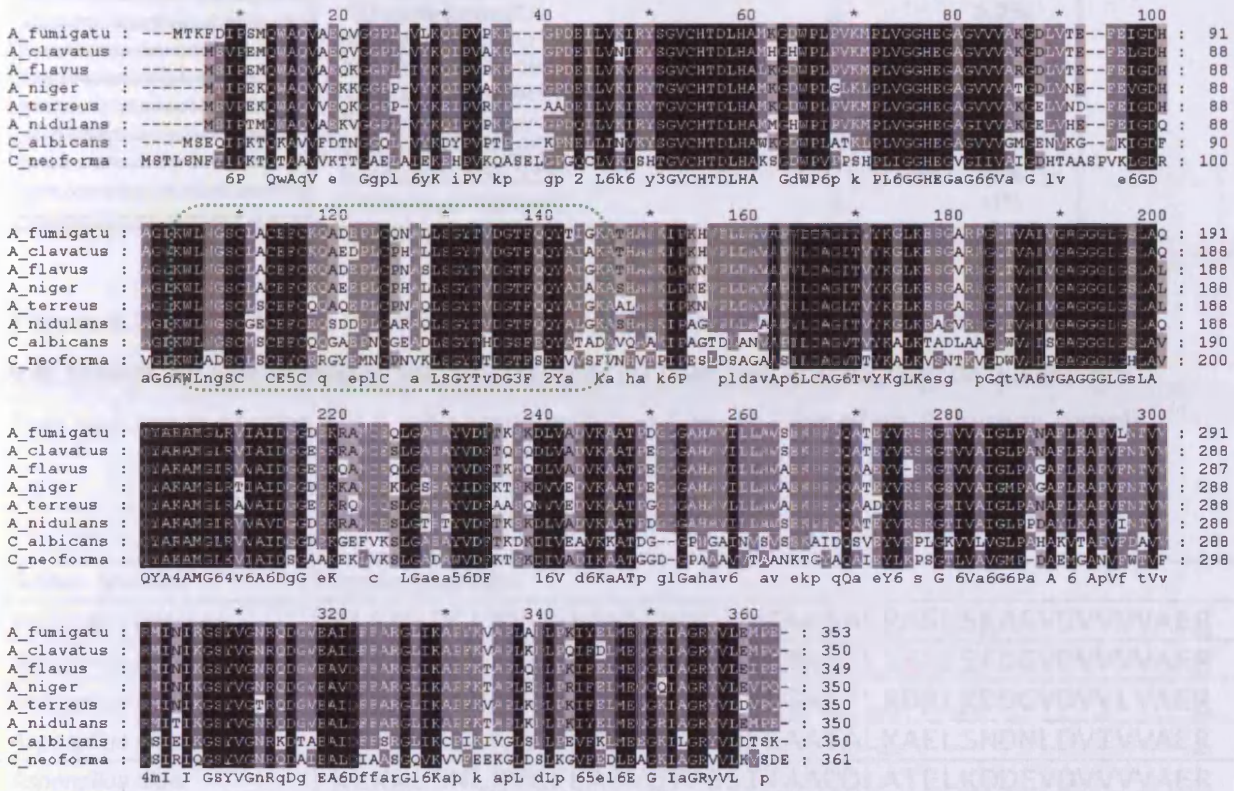


Figure 4.23. Multiple alignment of alcohol dehydrogenase sequences showing conserved and variable regions between test species. A region spanning tryptic peptides diagnostic for classifying Aspergilli to species (Table 4.6) is outlined in green.

Catalase. Catalase (44 kDa), was not an abundant protein, based on the number of peptides detected and the low sequence coverage (Figure 4.24), but was detected in a number of samples and yielded peptides that could be used to classify Aspergilli. Tryptic peptides 'VGFLASVETPASIEAASELS**K**QLSEDGVDVVVVAE**R**' from section Flavi and the peptides from corresponding regions of other Aspergilli are often detected and are diagnostic for classification of the fungi, Table 4.7. The multiple sequence alignment of complete catalase amino acid sequences from the test species and the region with tryptic peptides diagnostic for species is presented in Figure 4.25.

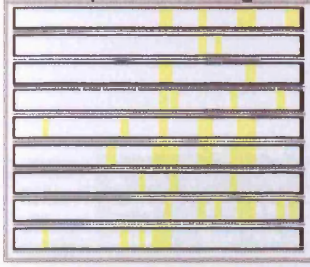
Sequence coverage	Sample	# Peptides	%Coverage
	01AflavusNRRL1957	6	13%
	13AparasiticusSU1	2	3.2%
	02AparasiticusSU1 (conidia)	3	5.4%
	15AflavusNRRL1957	4	7.4%
	12AparasiticusNRRL502	7	15%
	A_parasiticus_myc01	10	20%
	13AoyzaeNRRL5590	3	4.8%
	09AflavusNRRL3357	7	14%
	03AterreusNRRL255	6	11%

Figure 4.24. LC/MS/MS peptide coverage for catalase from *A. oryzae* across several test strains. In this example, most of the spectra matched to *A. oryzae* catalase were from species in section Flavi, with some matches from spectra of an *A. terreus* sample.

Genus species	tryptic peptides
Aspergillus clavatus	**K**L**KK**LDGL**K**VGFFASVQHASSLDAASAL**R**ASLS**K**AGVDVVVVAE**R**
Aspergillus flavus/oryzae	**K**L**KK**LDGL**K**VGFLASVETPASIEAASELS**K**QLSEDGVDVVVVAE**R**
Aspergillus fumigatus	**K**L**KK**LDGL**K**VGVLGSVQHPGSVEGASTL**R**D**R**L**K**DDGVDVVLVAE**R**
Aspergillus nidulans	**R**LQSLAGL**K**IAVLASVDAEESFSAATAL**K**AELSNDNLDVIVVAE**R**
Aspergillus niger	**K**L**KK**LSNL**R**VGFLASVQTPSSITAAQDLATEL**K**DDEVDVVVVAE**R**
Aspergillus terreus	**K**LL**R**LDGL**K**VAVLGSVDVPDSLAA**R**QTIASQLAGENVDVVTVAE**R**

Table 4.7. Region within catalase with tryptic peptides diagnostic for identification of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**. In this region, there are a number of differences in amino acids and an altered cleavage site (marked in red).

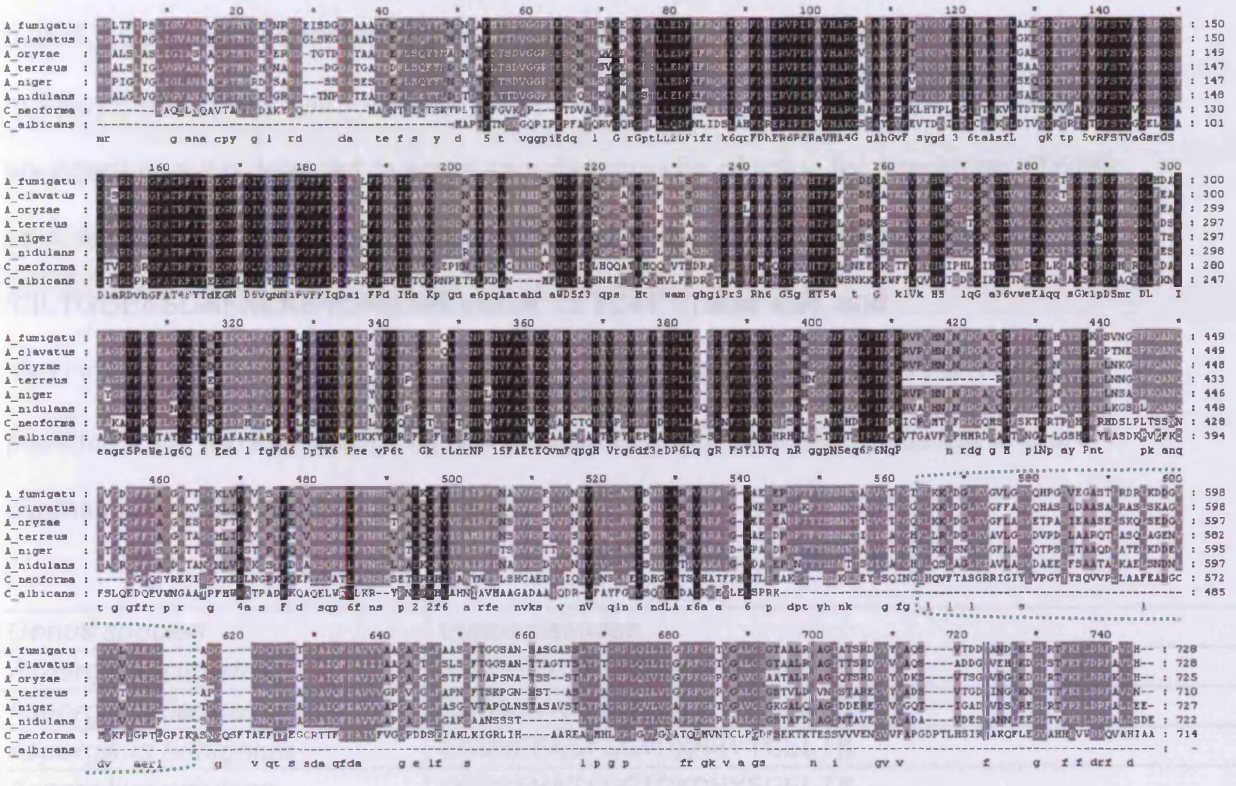


Figure 4.25. Multiple sequence alignment of catalase sequences showing conserved and variable regions between species. A region spanning tryptic peptides diagnostic for classifying Aspergilli to species (Table 4.7) is outlined in green.

Less abundant proteins. Aspartic endopeptidase (~43 kDa), translationally-controlled tumor protein (TCTP, 20kDa), and cobalamine-independent methionine synthase (CIMS, ~86 kDa, also referred to as 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase) were additional proteins that, while not always abundant or even detected in some samples, may be of value for speciation. Tryptic peptides 'EGDDSVATFGGVD**K**DHYTGELV**K**' of aspartic endopeptidase (Table 4.8), 'DILTGDEIISDAFNL**K**EVDNILWEVDC**R**' of TCTP (Table 4.9), and 'GQTVDPIT**K**INDLLPVYVELLQ**K**' (Table 4.10) of CIMS from section Flavi and the peptides from corresponding regions of other Aspergilli are often detected and are diagnostic for classification of the fungi.

Genus species	tryptic peptides
Aspergillus clavatus	EGDSSVATFGGIDKDHFTGELTK
Aspergillus flavus/oryzae	EGDDSVATFGGVDKDHYTGELVK
Aspergillus fumigatus	EGDNSEASFGGVDKNHYTGELTK
Aspergillus nidulans	DGDSSVATFGGIDKDHYEGELIK
Aspergillus niger	EGDESVATFGGVDKDHYTGELIK
Aspergillus terreus	EGDESVATFGGVDKSHYTGELIK

Table 4.8. Region within aspartic endopeptidase with tryptic peptides diagnostic for identification of Aspergilli. Cleavage sites at lysine residues are in **BOLD**.

Genus species	tryptic peptides
Aspergillus clavatus	MIIYKDIISGDEVLADTFNIKTVDGVLYECDCR
Aspergillus flavus/oryzae	MIIYKDILTGDEIISDAFNLKEVDNILWEVDCR
Aspergillus fumigatus	MIIYKDIISGDEVLSDNFKIKEVDGVLYECDCR
Aspergillus nidulans	MIIYKDIISGDEVLSDTYNIKTVDGVLYECDCR
Aspergillus niger	MIIYTDIVSGDEVLSDTFKIQEDSDSKLLWTCDCR
Aspergillus terreus	MIIYKDIISGDEVLSDTFNIKTVDNVFYECDCR

Table 4.9. Region within translationally-controlled tumour protein with tryptic peptides diagnostic for identification of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**.

Genus species	tryptic peptides
Aspergillus clavatus	GQTVDPIDKINDLLPLYVDLLAKLK
Aspergillus flavus/oryzae	GQTVDPITKINDLLPVYVELLQKLK
Aspergillus fumigatus	GQKIDPIDKINDLVPVYVDLLAQLK
Aspergillus nidulans	GQTLDPISKIDELLPLYVELLTKLK
Aspergillus niger	GQSVDPITKIEELLPVYVELLQKLK
Aspergillus terreus	GQTLDPIDKINELLPVYVELLTKLK

Table 4.10. Region within cobalamine-independent methionine synthase with tryptic peptides diagnostic for speciation of Aspergilli. Cleavage sites at lysine residues are in **BOLD**.

Summary. The objectives of this phase of the study was to characterise a cross section of reference strains using LC/MS and LC/MS/MS profiling techniques in order to find specific small molecules, peptides or biochemical profiles that can be used to classify the species. A targeted analytical method was developed for mycotoxin analysis and may be useful for confirmation of identity, but was of little use as a primary characteristic for classification of Aspergilli. The methodology employed for small molecule profiling by LC/MS yielded much richer data sets than GC/MS. However, the LC/MS data was less reproducible and would have required significantly more development to refine the technique for quantitative comparisons. Qualitative review of the profiling data yielded a small molecule and a peptide-like compound that may be of use for discriminating between species or strains within section Flavi, but these compounds have yet to be identified. A method was developed for preparing protein extracts of fungal biomass and analyzing peptides from tryptic digests of those extracts by LC/MS/MS. Protein profiling by LC/MS/MS yielded matches to a large number of fungal proteins and was very effective for identifying many of the *Aspergillus* reference strains from pure lab cultures. Additional method development was required to adapt the methods employed for the reference strains to allow identification of field isolates and *Aspergillus* associated with host tissue and this is presented with the case studies in Chapter 5.

Chapter 5: Results of Case Studies for Detection and Identification of *Aspergillus*

Aspergillus species can be readily isolated from soil or biological samples and are most often identified by their distinctive morphological characteristics. In Chapter 4, biochemical profiling analyses of *Aspergillus* reference strains demonstrated that, under carefully controlled culture conditions, protein profiling was effective for classifying many Aspergilli to the species level. In this phase of the study, protein profiling was applied to detecting and identifying *Aspergillus flavus* under a wider range of conditions, in scenarios typically encountered with laboratory and field studies of *Aspergillus*. Specifically, protein profiling was applied to identification of laboratory and field isolates of fungi and to detecting the presence of *Aspergillus flavus* in maize kernels. Also, supplemental analyses were conducted on additional reference strains from the USDA that include two strains of *Penicillium chrysogenum* along with strains of *Aspergillus carbonarius*, *Aspergillus niger*, *Petromyces alliaceous* and *Paecilomyces lilacinus*. These supplementary analyses were conducted to complement the previous protein profiling analyses with additional closely related species and add replicate reference strains of *P. chrysogenum*, *A. carbonarius* and *A. niger* (previously *A. carbonarius* and *A. niger* were each represented by only one reference strain).

Culture and morphology of isolates. Clayton 270, Sandhills 174, and Plymouth 98 were isolated and tentatively identified as *Aspergillus flavus* based on macroscopic characteristics of growth habit and colour and microscopic characteristics of the

conidiophores and conidia. All have *Aspergillus*-like conidiophores with characteristic golden yellow to olive green conidia (Figures 5.1).

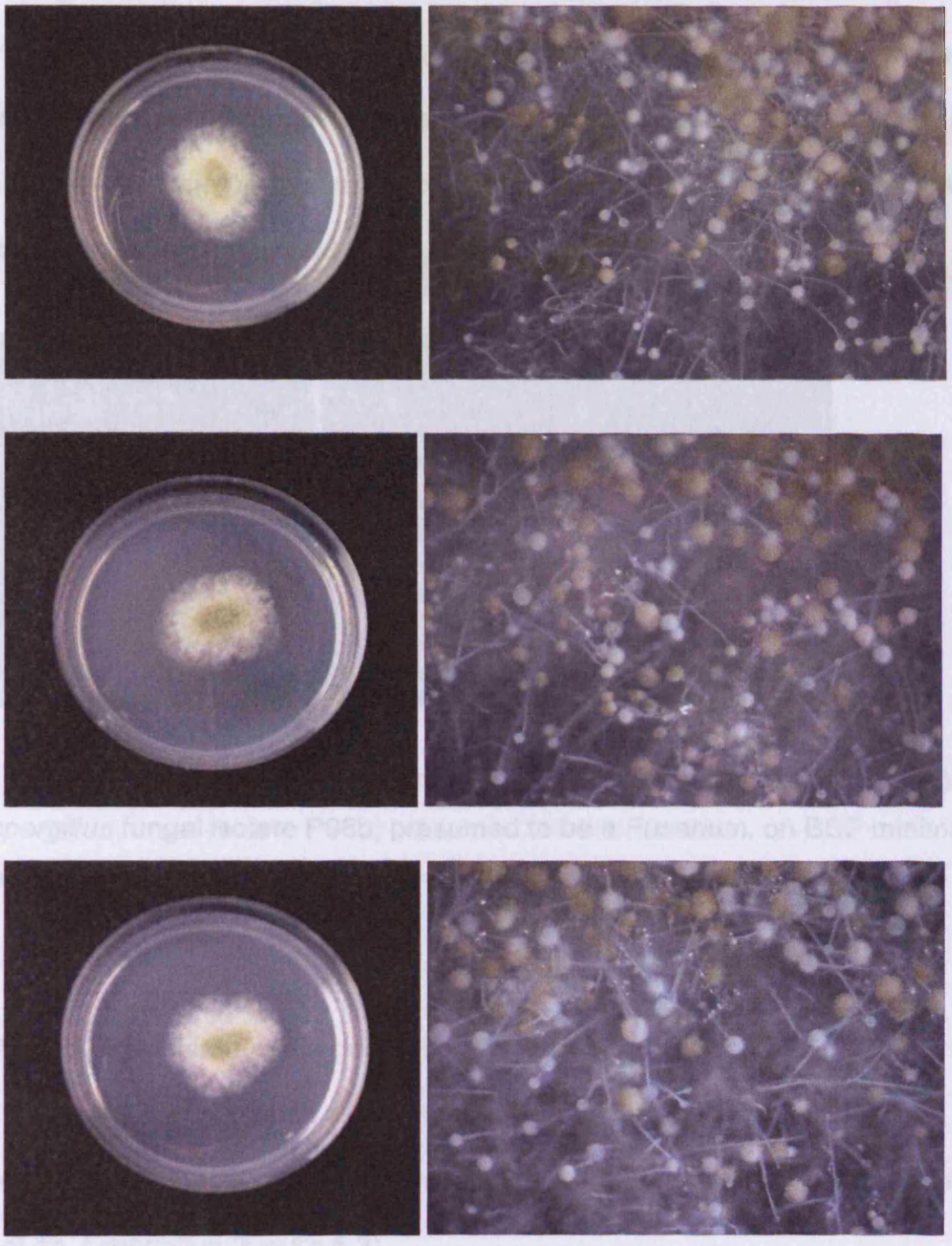


Figure 5.1. Colour and growth habit on BCP minimal medium, day 4 (left), and conidiophores (right) of presumed *Aspergillus flavus* isolates Clayton 270 (A), Sandhills 174 (B), and Plymouth 98 (C).

Isolate P98b had white, finely branched structures close to the surface of the growth medium that gained a pink cast with age and was tentatively identified as a *Fusarium* species (Figure 5.2), most likely *Fusarium verticilliodes* (syn. *F. moniliforme*), but possibly *Fusarium graminearum*. Although this isolate lacked distinctive reproductive structures, *Fusarium* is a fungus commonly isolated from maize (*F. verticilliodes* is the predominant species) and the presence of pink or red mold is very characteristic for *Fusarium* infection (Munkvold 2003).

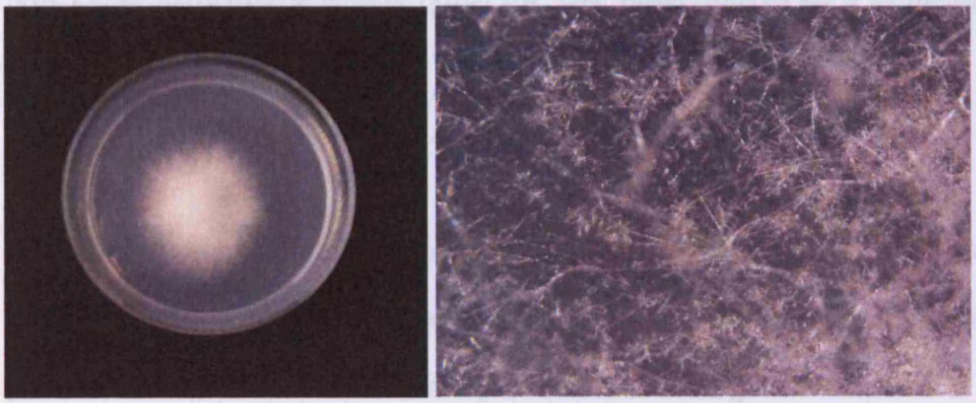


Figure 5.2. Colour and growth habit (left) and microscopic characteristics (right) of non-*Aspergillus* fungal isolate P98b, presumed to be a *Fusarium*, on BCP minimal medium, day 4.

The GAPLAB_RH isolate had a general growth habit and grey/green colour that was similar to some Aspergilli, such as *Aspergillus fumigatus* or *Aspergillus nidulans*, but had branched, broom-like conidiophores of a *Penicilllium* rather than defined vesicles typical of an *Aspergillus* (Figure 5.3).

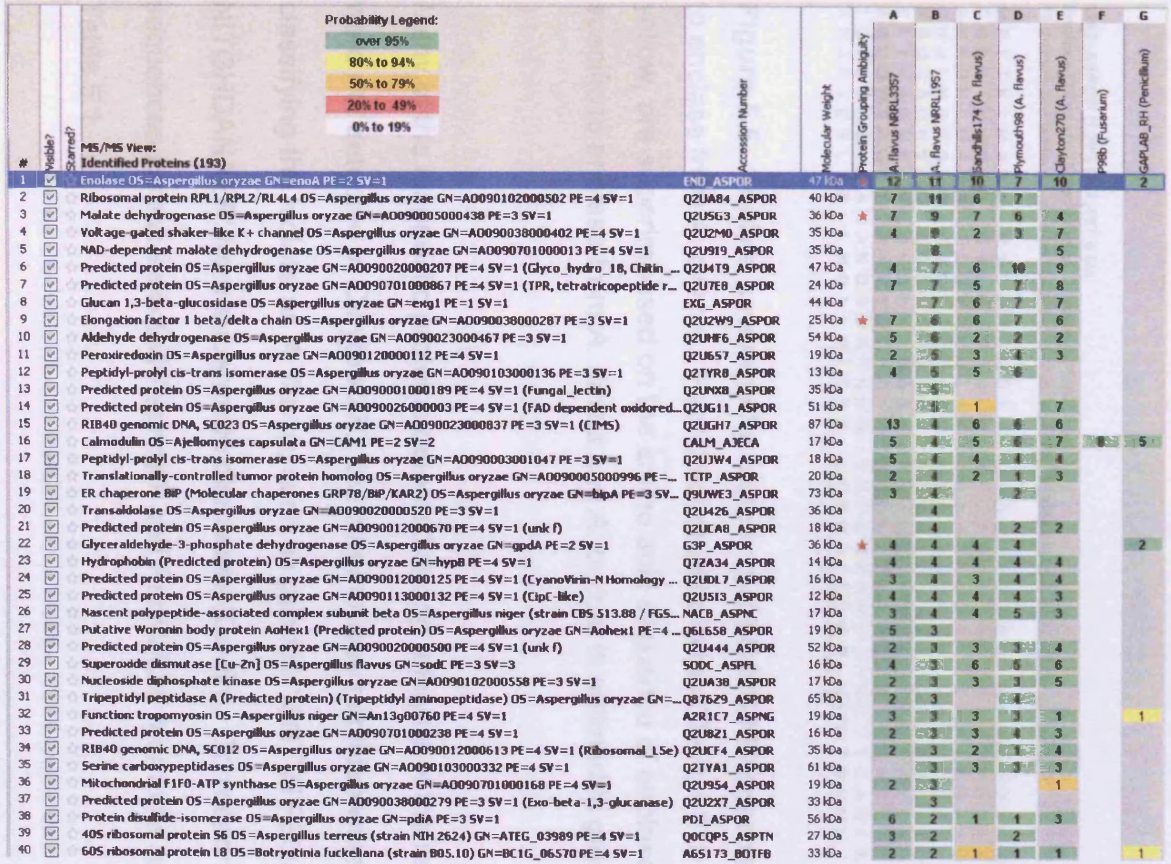
Figure 5.3. Colour and growth habit (left) and microscopic characteristics (right) of non-*Aspergillus* fungal isolate GAPLAB_RH, presumed to be a *Penicillium*, on BCP minimal medium, day 4.

Aflatoxin. Analysis of growth media and infected maize kernels by LC/MS/MS revealed that the reference strain of *Aspergillus flavus*, NRRL 3357, consistently produced aflatoxin under laboratory conditions when grown on minimal medium or PDA. One of the three *Aspergillus* isolates, Clayton 270, did not produce aflatoxins under any of the conditions tested, and so aflatoxin analysis neither supported nor contraindicated the preliminary identification of this isolate as *Aspergillus flavus*. The Sandhill 174 and Plymouth 98 isolates did not produce significant amounts of aflatoxins in laboratory culture, but did yield aflatoxins B1 and B2 in infected maize kernels (data not shown), which supports the preliminary identification of these isolates as *Aspergillus flavus*. As expected, no aflatoxins were detected in growth medium of the non-*Aspergillus* isolates (not tested in maize kernels).

Protein profiling of isolates. Profiling of proteins from biomass scraped off agar-solidified medium yielded approximately 10,000 MS/MS spectra per sample from, on average, 27 mg of starting material. When searched versus a fungal subset of UniProt Knowledgebase (UniProtKB, Universal Protein Resource), a total of 183

proteins were identified from the fungal isolates (Figure 5.4). Although a fungal subset of the National Center for Biotechnology Information (NCBI) protein database was used with the reference strains, UniProt was selected as the reference database for this set because it is a more compact database (speeds processing) and tends to have more complete annotation which reduces the need for manual annotation.

Protein profiling supported the initial morphological identification of the *Aspergillus* field isolates as *Aspergillus flavus* (Figure 5.4). The predominant protein identifications for the *Aspergillus flavus* field isolates were to *Aspergillus flavus* or *Aspergillus oryzae* database entries and the types of proteins identified were consistent with the reference strains of *Aspergillus flavus* analyzed concurrently with the isolates. There were only a few low quality matches of spectra from P98b (*Fusarium*) and GAPLAB_RH (*Penicillium*) to *Aspergillus* proteins. The modified sampling (scraping fungal material from agar) and the simplified protein preparation yielded a different cross section of proteins than in the previous experiments with the reference strains (Chapter 4) in that many of the major proteins, such as enolase, GAPDH, ribosomal proteins and chaperonins, are still detected and have good sequence coverage, but their relative abundances have changed. Also, some of the proteins useful for classification, such as alcohol dehydrogenase and pyruvate decarboxylase, are less abundant in this data set and lack the coverage required for use in identification. However, identifications could be made based on selected proteins/peptides as previously determined along with examination of additional proteins/peptides detected in this study.



Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

#	Visible?	Starred?	MS/MS View: Identified Proteins (193)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	A (A. flavus NRRL3357)	B (A. flavus NRRL1957)	C (Sandhills174 (A. flavus))	D (Plymouth98 (A. flavus))	E (Clayton270 (A. flavus))	F (P98b (Fusarium))	G (GAPLAB_RH (Penicillium))
1	☑	☐	Enolase OS=Aspergillus oryzae GN=enoA PE=2 SV=1	ENO_ASPOR	47 kDa	*	12	11	10	7	10		2
2	☑	☐	Ribosomal protein RPL1/RPL2/RL4L4 OS=Aspergillus oryzae GN=AO090102000502 PE=4 SV=1	Q2UA84_ASPOR	40 kDa		7	11	6	7			
3	☑	☐	Malate dehydrogenase OS=Aspergillus oryzae GN=AO090005000438 PE=3 SV=1	Q2U5G3_ASPOR	36 kDa	*	7	9	7	6	4		
4	☑	☐	Voltage-gated shaker-like K+ channel OS=Aspergillus oryzae GN=AO090038000402 PE=4 SV=1	Q2U2M0_ASPOR	35 kDa		4	8	2	3	7		
5	☑	☐	NAD-dependent malate dehydrogenase OS=Aspergillus oryzae GN=AO090701000013 PE=4 SV=1	Q2U9I9_ASPOR	35 kDa			8			5		
6	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090020000207 PE=4 SV=1 (Glyco_hydro_18, Chitin_...	Q2U4T9_ASPOR	47 kDa		4	7	6	10	9		
7	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090701000867 PE=4 SV=1 (TPR, tetratricopeptide r...	Q2U7E8_ASPOR	24 kDa		7	7	5	7	8		
8	☑	☐	Glucan 1,3-beta-glucosidase OS=Aspergillus oryzae GN=exg1 PE=1 SV=1	EXG_ASPOR	44 kDa			7	6	7	7		
9	☑	☐	Elongation factor 1 beta/delta chain OS=Aspergillus oryzae GN=AO090038000287 PE=3 SV=1	Q2U2W9_ASPOR	25 kDa	*	7	6	6	7	6		
10	☑	☐	Aldehyde dehydrogenase OS=Aspergillus oryzae GN=AO090023000467 PE=3 SV=1	Q2UHF6_ASPOR	54 kDa		5	6	2	2	2		
11	☑	☐	Peroxiredoxin OS=Aspergillus oryzae GN=AO090120000112 PE=4 SV=1	Q2U6S7_ASPOR	19 kDa		2	5	3	4	3		
12	☑	☐	Peptidyl-prolyl cis-trans isomerase OS=Aspergillus oryzae GN=AO090103000136 PE=3 SV=1	Q2TYR8_ASPOR	13 kDa		4	5	5	6			
13	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090001000189 PE=4 SV=1 (Fungal_lectin)	Q2UNX8_ASPOR	35 kDa			5					
14	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090026000003 PE=4 SV=1 (FAD dependent oxidored...	Q2UG11_ASPOR	51 kDa			1	1		7		
15	☑	☐	RIB40 genomic DNA, SC023 OS=Aspergillus oryzae GN=AO090023000837 PE=3 SV=1 (CIMS)	Q2UGH7_ASPOR	87 kDa		13	4	6	6	6		
16	☑	☐	Calmodulin OS=Ajellomyces capsulata GN=CAM1 PE=2 SV=2	CALM_AJECA	17 kDa		5	4	5	6	7	6	5
17	☑	☐	Peptidyl-prolyl cis-trans isomerase OS=Aspergillus oryzae GN=AO090003001047 PE=3 SV=1	Q2UJW4_ASPOR	18 kDa		5	4	4	4	4		
18	☑	☐	Translationally-controlled tumor protein homolog OS=Aspergillus oryzae GN=AO090005000996 PE=...	TCTP_ASPOR	20 kDa		2	4	2	1	3		
19	☑	☐	ER chaperone BiP (Molecular chaperones GRP78/BiP/KAR2) OS=Aspergillus oryzae GN=bipA PE=3 SV...	Q9UWE3_ASPOR	73 kDa		3	4		2			
20	☑	☐	Transaldolase OS=Aspergillus oryzae GN=AO090020000520 PE=3 SV=1	Q2U426_ASPOR	36 kDa			4					
21	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090012000670 PE=4 SV=1 (unk f)	Q2UCA8_ASPOR	18 kDa			4		2	2		
22	☑	☐	Glyceraldehyde-3-phosphate dehydrogenase OS=Aspergillus oryzae GN=gpdA PE=2 SV=1	G3P_ASPOR	36 kDa	*	4	4	4	4			2
23	☑	☐	Hydrophobin (Predicted protein) OS=Aspergillus oryzae GN=hypB PE=4 SV=1	Q7ZA34_ASPOR	14 kDa		4	4	4	4	4		
24	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090012000125 PE=4 SV=1 (CyanoVirin-N Homology ...	Q2UDL7_ASPOR	16 kDa		3	4	3	4	4		
25	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090113000132 PE=4 SV=1 (CipC-like)	Q2U5I3_ASPOR	12 kDa		4	4	4	4	3		
26	☑	☐	Nascent polypeptide-associated complex subunit beta OS=Aspergillus niger (strain CBS 513.88 / FGS...	NACB_ASPNC	17 kDa		3	4	4	5	3		
27	☑	☐	Putative Woronin body protein AoHex1 (Predicted protein) OS=Aspergillus oryzae GN=AoHex1 PE=4 ...	Q6L658_ASPOR	19 kDa		5	3					
28	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090020000500 PE=4 SV=1 (unk f)	Q2U444_ASPOR	52 kDa		2	3	3	3	4		
29	☑	☐	Superoxide dismutase [Cu-Zn] OS=Aspergillus flavus GN=sodC PE=3 SV=3	SODC_ASPFL	16 kDa		4	3	6	5	6		
30	☑	☐	Nucleoside diphosphate kinase OS=Aspergillus oryzae GN=AO090102000558 PE=3 SV=1	Q2UA38_ASPOR	17 kDa		2	3	3	3	5		
31	☑	☐	Tripeptidyl peptidase A (Predicted protein) (Tripeptidyl aminopeptidase) OS=Aspergillus oryzae GN=...	Q87629_ASPOR	65 kDa		2	3		4			
32	☑	☐	Function: tropomyosin OS=Aspergillus niger GN=An13g00760 PE=4 SV=1	A2R1C7_ASPNG	19 kDa		3	3	3	3	1		1
33	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090701000238 PE=4 SV=1	Q2UB21_ASPOR	16 kDa		2	2	3	4	3		
34	☑	☐	RIB40 genomic DNA, SC012 OS=Aspergillus oryzae GN=AO090012000613 PE=4 SV=1 (Ribosomal_L5e)	Q2UCF4_ASPOR	35 kDa		2	3	2	1	4		
35	☑	☐	Serine carboxypeptidases OS=Aspergillus oryzae GN=AO090103000332 PE=4 SV=1	Q2TYA1_ASPOR	61 kDa			3	3	3	3		
36	☑	☐	Mitochondrial F1F0-ATP synthase OS=Aspergillus oryzae GN=AO090701000168 PE=4 SV=1	Q2U954_ASPOR	19 kDa		2	2			1		
37	☑	☐	Predicted protein OS=Aspergillus oryzae GN=AO090038000279 PE=3 SV=1 (Exo-beta-1,3-glucanase)	Q2U2X7_ASPOR	33 kDa			3					
38	☑	☐	Protein disulfide-isomerase OS=Aspergillus oryzae GN=pdiA PE=3 SV=1	PDI_ASPOR	56 kDa		6	2	1	1	3		
39	☑	☐	40S ribosomal protein S6 OS=Aspergillus terreus (strain NIH 2624) GN=ATEG_03989 PE=4 SV=1	Q0CQP5_ASPTN	27 kDa		3	2		2			
40	☑	☐	60S ribosomal protein L8 OS=Botryotinia fuckeliana (strain B05.10) GN=BCIG_06570 PE=4 SV=1	A6S173_BOTFB	33 kDa		2	2	1	1	1		1

Figure 5.4. An example view of a Scaffold™ results table compiled from LC/MS/MS analyses of two *Aspergillus flavus* reference strains and five fungal isolates. Results have been sorted on type strain *A. flavus* NRRL1957 (column B) and the number of unique peptides detected. Note that major proteins identified for *A. flavus* isolates match *A. flavus/oryzae* entries, but *Fusarium* and *Penicillium* do not.

Clayton 270 grouped with *A. flavus* based on the large number of good quality *A. flavus/oryzae* protein matches and to a high quality match with the tryptic peptide 'IEEELGNNAIYAGE**K**' specific to enolase from *A. flavus* and *A. oryzae*. The coverage for enolase is presented in Figure 5.5.

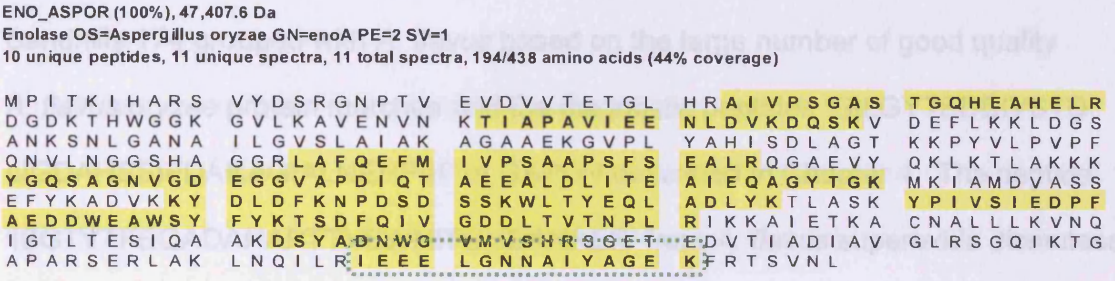


Figure 5.5. Peptide coverage for matching of LC/MS/MS spectra from Clayton 270 to enolase from *A. oryzae*. Peptides detected in the Clayton 270 sample, shown in yellow, are superimposed on the amino acid sequence of enolase. A tryptic peptide specific to enolase from *A. flavus* and *A. oryzae* is outlined in green.

Many of the diagnostic proteins detected during analysis of the reference strains were not detected in Clayton 270, but other unique peptides were present to support classifying the isolate as *A. flavus*, such as peptide 'ISGTVTFEQADANAPTTVSW NITGHDANAE**R**' from superoxide dismutase [Cu-Zn]. Corresponding amino acid sequences across selected Aspergilli and *Penicillium marneffei* are presented in Table 5.1.

Genus species	tryptic peptides
Aspergillus clavatus	GDS**K**VSGTVTFEQADENSLTTVSWNITGHDANA**K**
Aspergillus flavus/oryzae	GDS**K**ISGTVTFEQADANAPTTVSWNITGHDANAE**R**
Aspergillus fumigatus	GDS**K**ITGTVTFEQADENSPTTVSWNI**K**GNDPNA**K**
Aspergillus nidulans	GDS**K**VSGTVTFEQADENSNTTVSWNITGNDPNAE**R**
Aspergillus niger	GDS**K**VSGTVTFEQANENTPTTISWNITGHDANAE**R**
Aspergillus terreus	GDS**K**VVSGTVTFEQADANSLTTISWNITGNDPNAE**R**
Penicillium marneffei	GDSNI**K**GTVTFEQADENSPTTISWNITGHDANAE**R**

Table 5.1. Region within superoxide dismutase [Cu-Zn] that spans tryptic peptides diagnostic for speciation of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**.

Sandhills 174 grouped with *A. flavus* based on the large number of good quality *A. flavus/oryzae* protein matches and the diagnostic peptides 'GILGYTEDDIVSTD LIGDAHSSIFDA**K**AGIALNEHFI**K**' of GAPDH described in Chapter 4. The peptide 'ISGTVTFEQADANAPTTVSWNITGHDANAE**R**' from *A. flavus* superoxide dismutase [Cu-Zn] was also detected in the Sandhills 174 sample. Although enolase was an abundant protein in the Sandhills sample, the species-diagnostic peptide 'IEEELGNNAIYAGE**K**' from enolase was not detected. However, other tryptic peptides from enolase grouped Sandhills 174 with Aspergilli and a small tryptic peptide 'ACNALLL**K**' from enolase narrowed the identification to enolase from *A. flavus/oryzae*. The sequence coverage for enolase is presented in Figure 5.6.

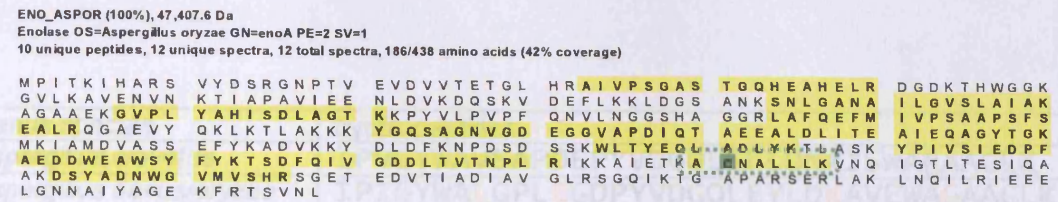


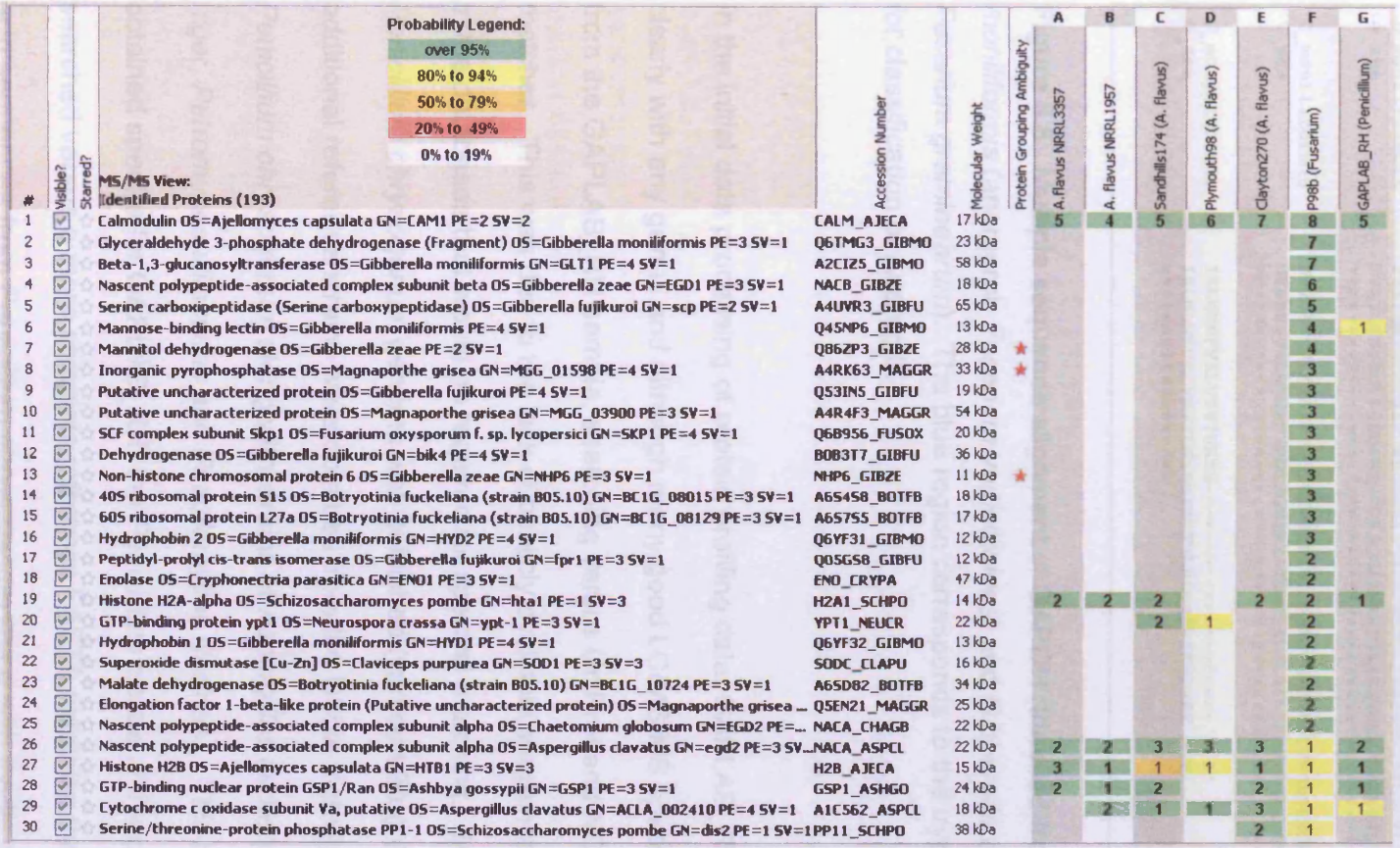
Figure 5.6. Peptide coverage for matching of LC/MS/MS spectra from Sandhills 174 to enolase from *A. oryzae*. Peptides detected in the Sandhills 174, shown in yellow, are superimposed on the amino acid sequence of enolase. Many spectra of enolase grouped Sandhills 174 with the Aspergilli and a match to the small peptide outlined in green, within the Aspergilli, was unique to *A. flavus/oryzae*.

Plymouth 98, like the other *A. flavus* isolates, grouped with *A. flavus* based on the large number of quality matches with *A. flavus/oryzae* protein matches, but many of the previously described diagnostic peptides were not detected. However, peptide 'VPTANVSVVDLTC**R**' from GAPDH does support classification as *A. flavus*. In addition, tryptic peptides 'IPIGYWALGPLEGDPYVDGQLEYLD**K**AVEWAGAAGL**K**' from glucan 1,3-beta-glucosidase are diagnostic for speciation of Aspergilli and were detected in Plymouth 98, as well as Clayton 270, Sandhills 174, and the type strain *A. flavus* NRRL 1957. This protein was notably absent from *A. flavus* NRRL 3357. Corresponding amino acid sequences of glucan 1,3-beta-glucosidase across selected Aspergilli are presented in Table 5.2. Tryptic peptides from this region, when detected, are particularly discriminative because there are a number of amino acid substitutions and some of those impact the cleavage sites for the trypsin digestion.

Genus species	tryptic peptides
Aspergillus clavatus	IPVGYWAVSAPDEPYVDGQLEFLDNAISWA**R**AAGL**K**
Aspergillus flavus/oryzae	IPIGYWALGPLEGDPYVDGQLEYLD**K**AVEWAGAAGL**K**
Aspergillus fumigatus	IPIGYWAVSSLPDEPYVDGQLEYLDNAISWA**R**EAGL**K**
Aspergillus nidulans	IPIGYWAAAPLDGEPYVSGQLEHLDNAVAWA**R**AHNL**K**
Aspergillus niger	IPIGYWAVAPIDGEPYVSGQIDYLDQAVTWA**R**AAGL**K**
Aspergillus terreus	IPIGYWAVEALPGDPYVDGQLEYLD**R**AIEWAGAAGL**K**

Table 5.2. Region of glucan 1,3-beta-glucosidase that spans tryptic peptides diagnostic for speciation of Aspergilli. Cleavage sites at arginine and lysine residues are in **BOLD**. Note that the sequence of *A. flavus/oryzae* differs from other species by several amino acids and a cleavage site (marked in red).

P98b grouped with *Fusarium* based on several good quality matches to proteins from that genus, particularly GAPDH (Figure 5.7), but protein profiling did not classify the isolate to the species level because the major proteins matched to both *Fusarium verticilliodes* and *Fusarium graminearum*. Searching versus a custom *Fusarium* protein database generated by merging protein databases from NCBI, the Broad Institute, and UniProtKB did not resolve ambiguous matching. This may be due to incomplete protein data for *Fusarium* in the databases or because *F. verticilliodes* and *F. graminearum* do not exhibit the same level of variation in amino acid sequences as Aspergilli. GAPDH is an example of a protein with a sequence that is useful for classifying Aspergilli, but may not work well for *Fusarium* because the sequences are so similar between *Fusarium verticillioides* and *Fusarium graminearum* (Figure 5.8). A protein profiling experiment with *Fusarium* reference strains would be required to determine which proteins may be useful for speciation of *Fusarium* isolates, but was beyond the scope of this study.



Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

#	Visible?	Starred?	MS/MS View: Identified Proteins (193)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	A A.flavus NRRL3357	B A. flavus NRRL1957	C Sandhills174 (A. flavus)	D Plymouth98 (A. flavus)	E Clayton270 (A. flavus)	F P98b (Fusarium)	G GAPLAB_RH (Penicillium)
1	☑		Calmodulin OS=Ajellomyces capsulata GN=CAM1 PE=2 SV=2	CALM_AJECA	17 kDa		5	4	5	6	7	8	5
2	☑		Glyceraldehyde 3-phosphate dehydrogenase (Fragment) OS=Gibberella moniliformis PE=3 SV=1	Q6TMG3_GIBMO	23 kDa							7	
3	☑		Beta-1,3-glucanosyltransferase OS=Gibberella moniliformis GN=GLT1 PE=4 SV=1	A2CIZ5_GIBMO	58 kDa							7	
4	☑		Nascent polypeptide-associated complex subunit beta OS=Gibberella zeae GN=EGD1 PE=3 SV=1	NACB_GIBZE	18 kDa							6	
5	☑		Serine carboxipeptidase (Serine carboxypeptidase) OS=Gibberella fujikuroi GN=scp PE=2 SV=1	A4UVR3_GIBFU	65 kDa							5	
6	☑		Mannose-binding lectin OS=Gibberella moniliformis PE=4 SV=1	Q45NP6_GIBMO	13 kDa							4	1
7	☑		Mannitol dehydrogenase OS=Gibberella zeae PE=2 SV=1	Q86ZP3_GIBZE	28 kDa	★						4	
8	☑		Inorganic pyrophosphatase OS=Magnaporthe grisea GN=MGG_01598 PE=4 SV=1	A4RK63_MAGGR	33 kDa	★						3	
9	☑		Putative uncharacterized protein OS=Gibberella fujikuroi PE=4 SV=1	Q53IN5_GIBFU	19 kDa							3	
10	☑		Putative uncharacterized protein OS=Magnaporthe grisea GN=MGG_03900 PE=3 SV=1	A4R4F3_MAGGR	54 kDa							3	
11	☑		SCF complex subunit Skp1 OS=Fusarium oxysporum f. sp. lycopersici GN=SKP1 PE=4 SV=1	Q6B956_FUSOX	20 kDa							3	
12	☑		Dehydrogenase OS=Gibberella fujikuroi GN=bik4 PE=4 SV=1	B0B3T7_GIBFU	36 kDa							3	
13	☑		Non-histone chromosomal protein 6 OS=Gibberella zeae GN=NHP6 PE=3 SV=1	NHP6_GIBZE	11 kDa	★						3	
14	☑		40S ribosomal protein S15 OS=Botryotinia fuckeliana (strain B05.10) GN=BC1G_08015 PE=3 SV=1	A6S4S8_BOTFB	18 kDa							3	
15	☑		60S ribosomal protein L27a OS=Botryotinia fuckeliana (strain B05.10) GN=BC1G_08129 PE=3 SV=1	A6S7S5_BOTFB	17 kDa							3	
16	☑		Hydrophobin 2 OS=Gibberella moniliformis GN=HYD2 PE=4 SV=1	Q6YF31_GIBMO	12 kDa							3	
17	☑		Peptidyl-prolyl cis-trans isomerase OS=Gibberella fujikuroi GN=fpr1 PE=3 SV=1	Q0SGS8_GIBFU	12 kDa							2	
18	☑		Enolase OS=Cryphonectria parasitica GN=ENO1 PE=3 SV=1	ENO_CRYPA	47 kDa							2	
19	☑		Histone H2A-alpha OS=Schizosaccharomyces pombe GN=hta1 PE=1 SV=3	H2A1_SCHPO	14 kDa		2	2	2		2	2	1
20	☑		GTP-binding protein ypt1 OS=Neurospora crassa GN=ypt-1 PE=3 SV=1	YPT1_NEUCR	22 kDa				2	1		2	
21	☑		Hydrophobin 1 OS=Gibberella moniliformis GN=HYD1 PE=4 SV=1	Q6YF32_GIBMO	13 kDa							2	
22	☑		Superoxide dismutase [Cu-Zn] OS=Claviceps purpurea GN=SOD1 PE=3 SV=3	SODC_CLAPU	16 kDa							2	
23	☑		Malate dehydrogenase OS=Botryotinia fuckeliana (strain B05.10) GN=BC1G_10724 PE=3 SV=1	A6SD82_BOTFB	34 kDa							2	
24	☑		Elongation factor 1-beta-like protein (Putative uncharacterized protein) OS=Magnaporthe grisea ...	Q5EN21_MAGGR	25 kDa							2	
25	☑		Nascent polypeptide-associated complex subunit alpha OS=Chaetomium globosum GN=EGD2 PE=...	NACA_CHAGB	22 kDa							2	
26	☑		Nascent polypeptide-associated complex subunit alpha OS=Aspergillus clavatus GN=egd2 PE=3 SV...	NACA_ASPCL	22 kDa		2	2	3	3	3	1	2
27	☑		Histone H2B OS=Ajellomyces capsulata GN=HTB1 PE=3 SV=3	H2B_AJECA	15 kDa		3	1	1	1	1	1	1
28	☑		GTP-binding nuclear protein GSP1/Ran OS=Ashbya gossypii GN=GSP1 PE=3 SV=1	GSP1_ASHGO	24 kDa		2	1	2		2	1	1
29	☑		Cytochrome c oxidase subunit Va, putative OS=Aspergillus clavatus GN=ACLA_002410 PE=4 SV=1	A1C562_ASPCL	18 kDa			2	1	1	3	1	1
30	☑		Serine/threonine-protein phosphatase PP1-1 OS=Schizosaccharomyces pombe GN=dis2 PE=1 SV=1	PP11_SCHPO	38 kDa						2	1	

Figure 5.7. An example view of a Scaffold™ results table compiled from LC/MS/MS analyses of two *Aspergillus flavus* reference strains and five fungal isolates. Results have been sorted on *Fusarium* isolate P98b (column F) and the number of unique peptides detected. Note that many of the major proteins identified match *Fusarium* or *Gibberella* proteins, but protein matching does not resolve organism classification to species.

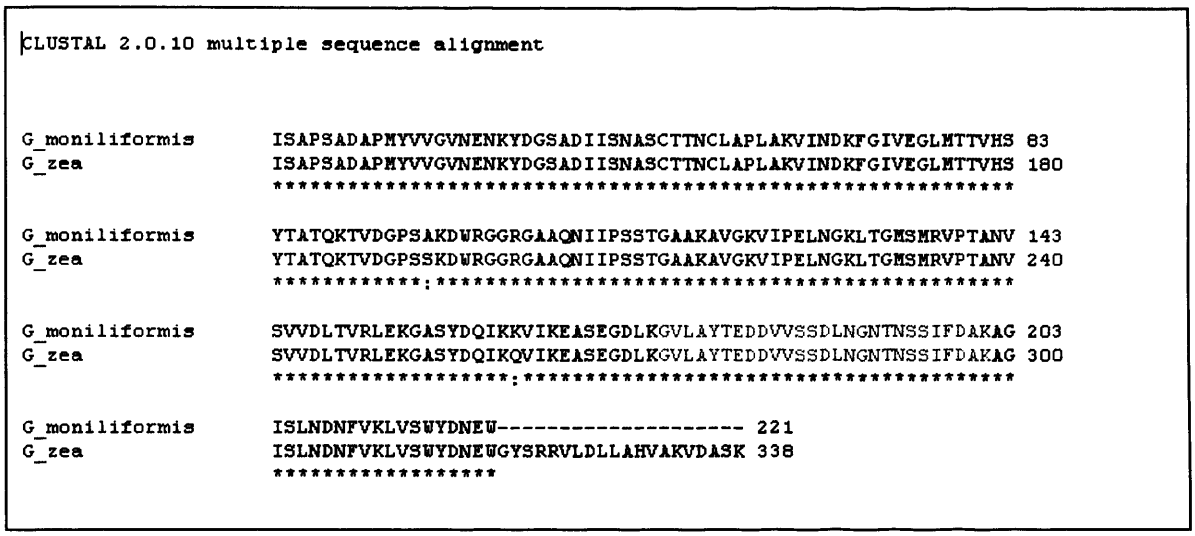
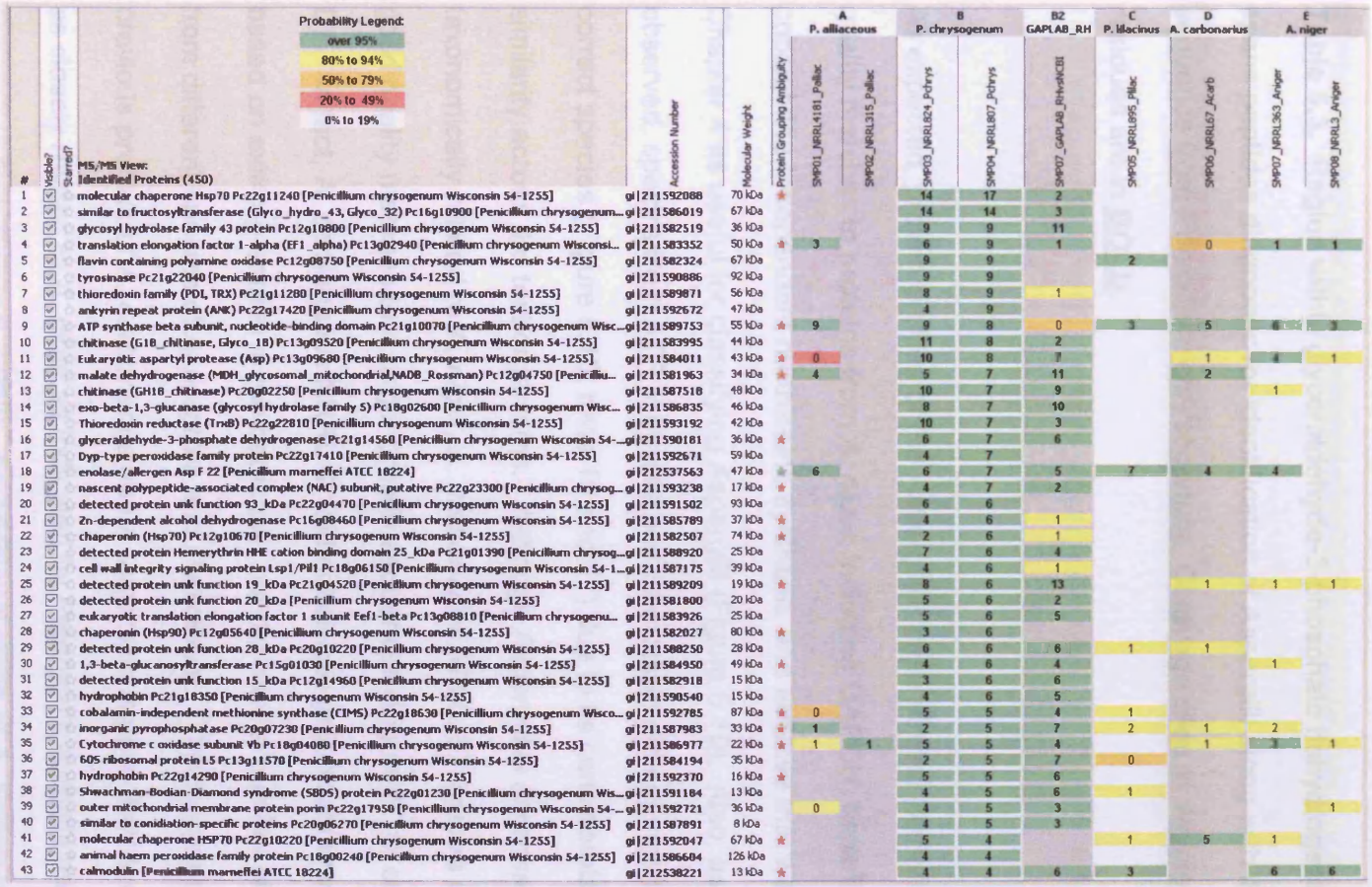


Figure 5.8. Multiple sequence alignment of GAPDH (trimmed)from *Gibberella moniliformis* (anamorph *Fusarium verticillioides*) and *Gibberella zea* (anamorph *Fusarium graminearum*). The blue region corresponds to the tryptic peptide useful for classification of Aspergilli.

In the initial data processing of protein profiling data, GAPLAB_RH did not group clearly with any genus and although many good LC/MS/MS spectra were generated from the GAPLAB_RH sample, searching versus UniProt and NCBI yielded very few matches. This was due to the lack of publicly available sequence data for *Penicillium* in the databases that would be required for spectral matching. Subsequently, the *Penicillium chrysogenum* genome was published (van den Berg *et al.* 2008) and additional reference strains were obtained from the USDA including two strains of *Penicillium chrysogenum* along with strains of *Aspergillus carbonarius*, *Aspergillus niger*, *Petromyces alliaceous* and *Paecilomyces lilacinus*. This allowed previously obtained spectra for GAPLAB_RH, the *Penicillium* isolate of the case studies, to be searched versus a more complete reference FASTA database that includes amino acid sequences from *Penicillium chrysogenum* along with tryptic digests of reference strains of *P. chrysogenum*.

Profiling of proteins from fungal biomass from the new reference strains yielded approximately 10,000 MS/MS spectra per sample and when searched versus a fungal subset of the National Center for Biotechnology Information (NCBI) protein database, the spectra from this dataset yielded matches to over 400 proteins. Reprocessing the previously collected MS/MS spectra for GAPLAB_RH (>9000 spectra) yielded matches to over 100 proteins almost exclusively to *P. chrysogenum*. In addition, the protein matches were consistent with the number and type of matches found for profiles of the two *P. chrysogenum* reference strains (Figure 5.9).

Of particular note, a tryptic peptide of GAPDH was detected in the *P. chrysogenum* type strain and in GAPLAB_RH that is comparable to a peptide diagnostic for classifying Aspergilli, 'GILGYTEDDIVSTDLIGDAHSSIFDA**K**' in *A. flavus* (Table 5.3). The tryptic peptide from *P. chrysogenum*, 'GILGYTEDQIVSTDLNGDE**R**' differs from that of Aspergilli by just a few amino acids and, as with *A. clavatus*, the amino acid sequence that follows this peptide lacks one of the cleavage sites present in a number of the Aspergillus species due to substitution of alanine for an arginine in the sequence.



Probability Legend:

over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

#	Visible?	Starred?	MS/MS View: Identified Proteins (450)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	A P. allaceous SMP01_NRRL4181_Pallac	A P. allaceous SMP02_NRRL315_Pallac	B P. chrysogenum SMP03_NRRL824_Pchrys	B P. chrysogenum SMP04_NRRL807_Pchrys	B2 GAPLAB_RH SMP07_GAPLAB_RHvsNCBI	C P. lilacinus SMP05_NRRL895_Pllac	D A. carbonarius SMP06_NRRL67_Acarb	E A. niger SMP07_NRRL363_Aniger	E A. niger SMP08_NRRL3_Aniger
1	☑	☐	molecular chaperone Hsp70 Pc22g11240 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211592088	70 kDa	*			14	17	2				
2	☑	☐	similar to fructosyltransferase (Glyco_hydro_43, Glyco_32) Pc16g10900 [Penicillium chrysogenum...	gi|211586019	67 kDa				14	14	3				
3	☑	☐	glycosyl hydrolase family 43 protein Pc12g10800 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211582519	36 kDa				9	9	11				
4	☑	☐	translation elongation factor 1-alpha (EF1_alpha) Pc13g02940 [Penicillium chrysogenum Wisconsi...	gi|211583352	50 kDa	*	3		6	9	1		0	1	1
5	☑	☐	flavin containing polyamine oxidase Pc12g08750 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211582324	67 kDa				9	9		2			
6	☑	☐	tyrosinase Pc21g22040 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211590886	92 kDa				9	9					
7	☑	☐	thioredoxin family (PDI, TRX) Pc21g11280 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211589871	56 kDa				8	9	1				
8	☑	☐	ankyrin repeat protein (ANK) Pc22g17420 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211592672	47 kDa				4	9					
9	☑	☐	ATP synthase beta subunit, nucleotide-binding domain Pc21g10070 [Penicillium chrysogenum Wisc...	gi|211589753	55 kDa	*	9		9	8	0	3	5	6	3
10	☑	☐	chitinase (G18_chitinase, Glyco_18) Pc13g09520 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211583995	44 kDa				11	8	2				
11	☑	☐	Eukaryotic aspartyl protease (Asp) Pc13g09680 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211584011	43 kDa	*	0		10	8	7		1	4	1
12	☑	☐	malate dehydrogenase (MDH_glycosomal_mitochondrial,NADB_Rossman) Pc12g04750 [Penicillu...	gi|211581963	34 kDa	*	4		5	7	11		2		
13	☑	☐	chitinase (GH18_chitinase) Pc20g02250 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211587518	48 kDa				10	7	9			1	
14	☑	☐	exo-beta-1,3-glucanase (glycosyl hydrolase family 5) Pc18g02600 [Penicillium chrysogenum Wisc...	gi|211586835	46 kDa				8	7	10				
15	☑	☐	Thioredoxin reductase (TrxB) Pc22g22810 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211593192	42 kDa				10	7	3				
16	☑	☐	glyceraldehyde-3-phosphate dehydrogenase Pc21g14560 [Penicillium chrysogenum Wisconsin 54-...	gi|211590181	36 kDa	*			6	7	6				
17	☑	☐	Dyp-type peroxidase family protein Pc22g17410 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211592671	59 kDa				4	7					
18	☑	☐	enolase/allergen Asp F 22 [Penicillium marneffei ATCC 18224]	gi|212537563	47 kDa	*	6		6	7	5	7	4	4	
19	☑	☐	nascent polypeptide-associated complex (NAC) subunit, putative Pc22g23300 [Penicillium chrysog...	gi|211593238	17 kDa	*			4	7	2				
20	☑	☐	detected protein unk function 93_kDa Pc22g04470 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211591502	93 kDa				6	6					
21	☑	☐	Zn-dependent alcohol dehydrogenase Pc16g08460 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211585789	37 kDa	*			4	6	1				
22	☑	☐	chaperonin (Hsp70) Pc12g10670 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211582507	74 kDa	*			2	6	1				
23	☑	☐	detected protein Hemerythrin HHE cation binding domain 25_kDa Pc21g01390 [Penicillium chrysog...	gi|211580920	25 kDa				7	6	4				
24	☑	☐	cell wall integrity signaling protein Lsp1/Pil1 Pc18g06150 [Penicillium chrysogenum Wisconsin 54-1...	gi|211587175	39 kDa				4	6	1				
25	☑	☐	detected protein unk function 19_kDa Pc21g04520 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211589209	19 kDa	*			8	6	13		1	1	1
26	☑	☐	detected protein unk function 20_kDa [Penicillium chrysogenum Wisconsin 54-1255]	gi|211581800	20 kDa				5	6	2				
27	☑	☐	eukaryotic translation elongation factor 1 subunit Eef1-beta Pc13g08810 [Penicillium chrysogenu...	gi|211583926	25 kDa				5	6	5				
28	☑	☐	chaperonin (Hsp90) Pc12g05640 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211582027	80 kDa	*			3	6					
29	☑	☐	detected protein unk function 28_kDa Pc20g10220 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211588250	28 kDa				6	6	6	1	1		
30	☑	☐	1,3-beta-glucanosyltransferase Pc15g01030 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211584950	49 kDa	*			4	6	4			1	
31	☑	☐	detected protein unk function 15_kDa Pc12g14960 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211582918	15 kDa				3	6	6				
32	☑	☐	hydrophobin Pc21g18350 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211590540	15 kDa				4	6	5				
33	☑	☐	cobalamin-independent methionine synthase (CIMS) Pc22g18630 [Penicillium chrysogenum Wisco...	gi|211592785	87 kDa	*	0		5	5	4	1			
34	☑	☐	inorganic pyrophosphatase Pc20g07230 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211587983	33 kDa	*	1		2	5	7	2	1	2	
35	☑	☐	Cytochrome c oxidase subunit Vb Pc18g04080 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211586977	22 kDa	*	1	1	5	5	4		1	1	1
36	☑	☐	60S ribosomal protein L5 Pc13g11570 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211584194	35 kDa				2	5	7	0			
37	☑	☐	hydrophobin Pc22g14290 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211592370	16 kDa	*			5	5	6				
38	☑	☐	Shwachman-Bodian-Diamond syndrome (SBDS) protein Pc22g01230 [Penicillium chrysogenum Wis...	gi|211591184	13 kDa				4	5	6	1			
39	☑	☐	outer mitochondrial membrane protein porin Pc22g17950 [Penicillium chrysogenum Wisconsin 54-...	gi|211592721	36 kDa		0		4	5	3				1
40	☑	☐	similar to conidiation-specific proteins Pc20g06270 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211587891	8 kDa				4	5	3				
41	☑	☐	molecular chaperone HSP70 Pc22g10220 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211592047	67 kDa	*			5	4		1	5	1	
42	☑	☐	animal haem peroxidase family protein Pc18g00240 [Penicillium chrysogenum Wisconsin 54-1255]	gi|211586684	126 kDa				4	4					
43	☑	☐	calmodulin [Penicillium marneffei ATCC 18224]	gi|212538221	13 kDa	*			4	4	6	3		6	6

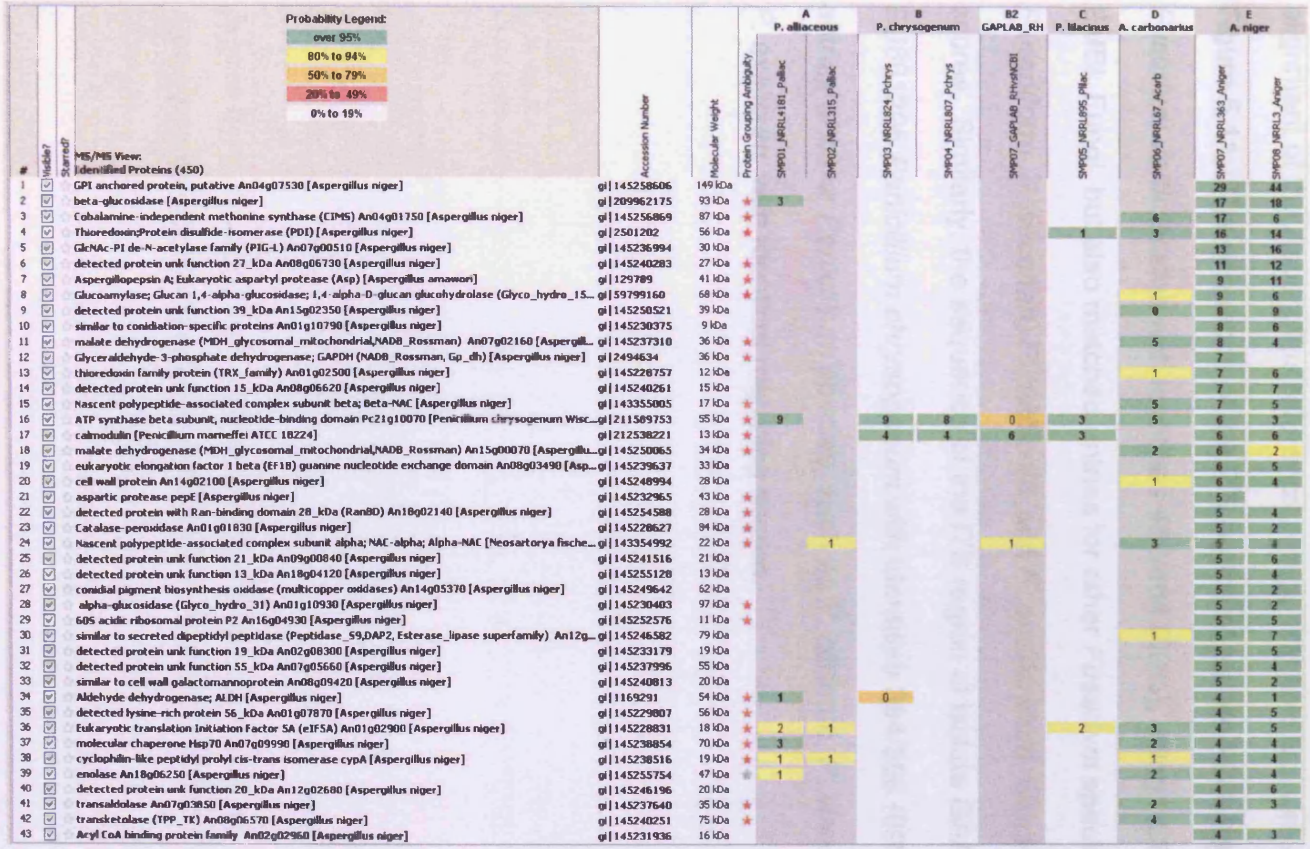
Figure 5.9. An example view of a Scaffold™ results table compiled from LC/MS/MS analyses of eight fungal reference strains and the GAPLAB_RH isolate. Results have been sorted on type strain *P. chrysogenum* NRRL 807 (column 4, category B) and the number of unique spectra detected. Note that major proteins identified for the GAPLAB_RH isolate match *Penicillium chrysogenum*, but other fungi do not.

Genus species	tryptic peptides
Aspergillus clavatus	GILGYTEDAIVSSDVNGDE**R**SSVFDAAAGISLNPNFV**K**
Aspergillus flavus/oryzae	GILGYTEDDIVSTDLIGDAHSSIFDA**K**AGIALNEHFI**K**
Aspergillus fumigatus	NILGYTEDDVVSSDLNGDE**R**SSIFDA**K**AGISLNPNFV**K**
Aspergillus nidulans	GILGYTEDDIVSTDLNGDT**R**SSIFDA**K**AGIALNSNFI**K**
Aspergillus niger	GILGYTEDDIVSSDLNGDDHSSIFDA**K**AGIALNSNFV**K**
Aspergillus terreus	GILGYTEDEVVSTDLNGDD**R**SSIFDA**K**AGIALNEHFV**K**
Penicillium chrysogenum	GILGYTEDQIVSTDLNGDE**R**SSVFDAAAGIALNANFI**K**

Table 5.3. Region within glyceraldehyde-3-phosphate dehydrogenase (GAPDH) that spans peptides diagnostic for identification of Aspergilli along with the comparable sequence from *Penicillium chrysogenum*. Cleavage sites at arginine and lysine residues are in **BOLD**.

As expected, protein profiles of the *Aspergillus niger* strains yielded many good quality matches to proteins from *A. niger*, including GAPDH, aspartic endopeptidase, cobalamine-independent methionine synthase, and enolase that were described in Chapter 4 as useful for classifying Aspergilli (Figure 5.10). Also, as previously observed, spectra from ATP synthase and calmodulin did not match samples to the correct species (Figure 5.10, rows 16 and 17) due to its considerable sequence similarity across the test organisms. Although *Aspergillus carbonarius* is taxonomically related to *A. niger*, *A. carbonarius* NRRL 67 spectra did not yield many good quality matches to *A. niger* protein sequences (Figure 5.10, column 7 category D). In fact, *A. carbonarius* NRRL 67 did not clearly group with any fungal species based on existing entries in the database, but did have a number of peptide matches from different species of *Aspergillus* and *Penicillium*. This is consistent with the previous profiles of *A. carbonarius* and *A. niger* and implies that *A. carbonarius* is not as closely related to *A. niger* as indicated by Peterson's analysis based on DNA sequences from four loci (2008). An additional strain of *A. carbonarius*, NRRL 346, did not grow readily on minimal medium and was excluded from these analyses.

As with *A. carbonarius*, protein profiling of *Petromyces alliaceous* did not clearly group with any fungal species based on existing entries in the database, but did have a number of peptide matches from different species of *Aspergillus* and *Penicillium*. *Paecilomyces lilacinus* did yield matches to a cross section of peptides from a number of fungi, but did not clearly group with any organism in the database (data not shown). This supports earlier observations that tryptic digests of organisms that lack substantial amino acid sequence data in the reference database yield few erroneous matches to proteins from other organisms. Also, as previously observed, ATP synthase appears to be a poor candidate protein for identifying fungi because there is enough shared sequence across species that spectra are often matched to ATP synthase from the wrong organism (Figure 5.10, line 16, ATP synthase from *P. chrysogenum*).



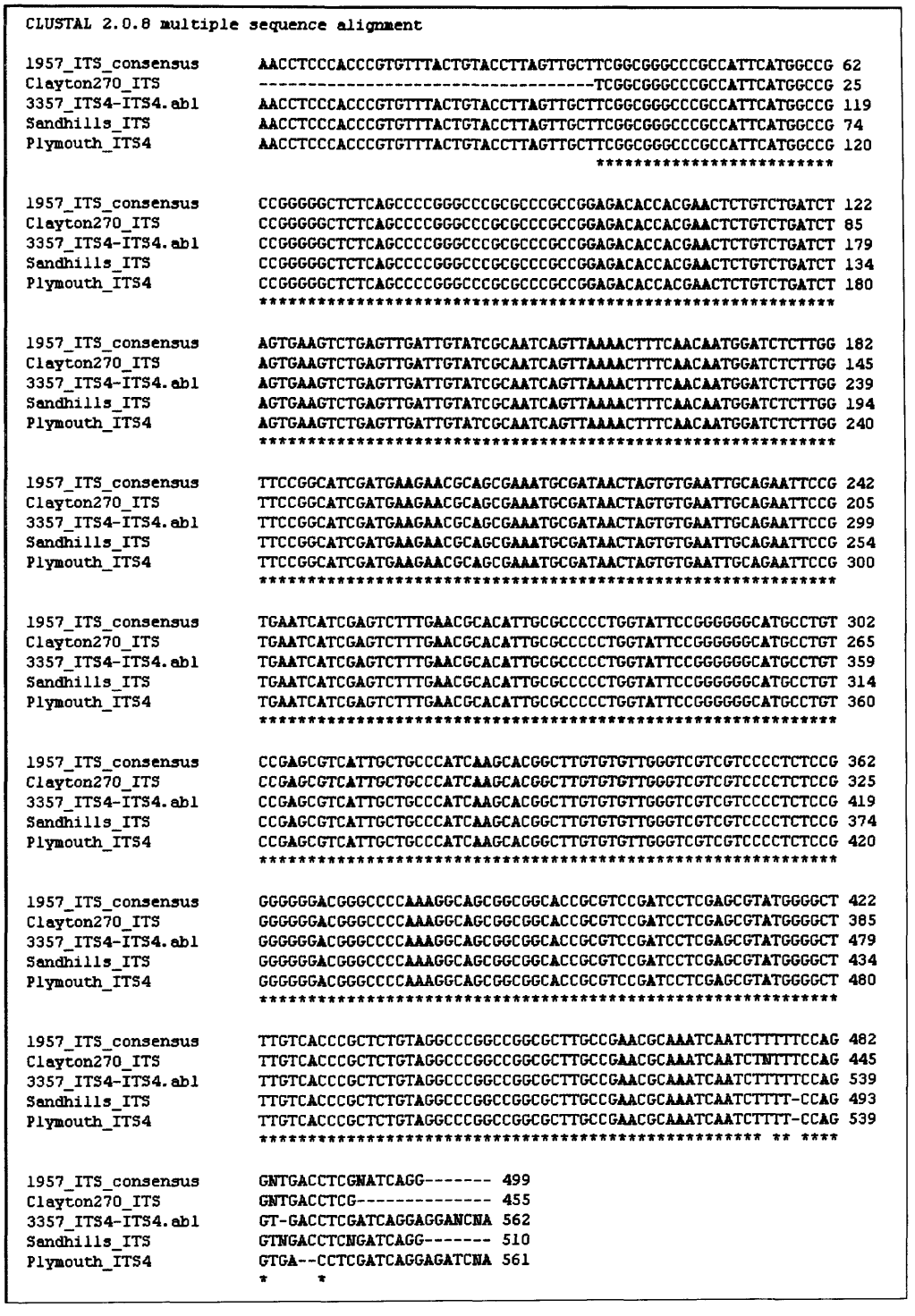
Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

MS/MS View:
Identified Proteins (450)

#	Visible?	Starred?		Accession Number	Molecular Weight	Protein Grouping Ambiguity	A P. allaceous SNP01_NRRL4181_Pallac	A P. allaceous SNP02_NRRL315_Pallac	B P. chrysogenum SNP03_NRRL824_Pchrys	B P. chrysogenum SNP04_NRRL807_Pchrys	B2 GAPLAB_RH SNP07_GAPLAB_RHvsNCBI	C P. lilacinus SNP05_NRRL895_Plilac	D A. carbonarius SNP06_NRRL67_Acarb	E A. niger SNP07_NRRL363_Aniger	E A. niger SNP08_NRRL3_Aniger
1			GPI anchored protein, putative An04g07530 [Aspergillus niger]	gi|145258606	149 kDa									29	44
2			beta-glucosidase [Aspergillus niger]	gi|209962175	93 kDa	*	3							17	18
3			Cobalamine-independent methonine synthase (CIMS) An04g01750 [Aspergillus niger]	gi|145256869	87 kDa	*							6	17	6
4			Thioredoxin;Protein disulfide-isomerase (PDI) [Aspergillus niger]	gi|2501202	56 kDa	*						1	3	16	14
5			GlcNAc-PI de-N-acetylase family (PIG-L) An07g00510 [Aspergillus niger]	gi|145236994	30 kDa									13	16
6			detected protein unk function 27_kDa An08g06730 [Aspergillus niger]	gi|145240283	27 kDa	*								11	12
7			Aspergillopepsin A; Eukaryotic aspartyl protease (Asp) [Aspergillus amawori]	gi|129789	41 kDa	*								9	11
8			Glucoamylase; Glucan 1,4-alpha-glucosidase; 1,4-alpha-D-glucan glucohydrolase (Glyco_hydro_15...	gi|59799160	68 kDa	*							1	9	6
9			detected protein unk function 39_kDa An15g02350 [Aspergillus niger]	gi|145250521	39 kDa								0	8	9
10			similar to conidiation-specific proteins An01g10790 [Aspergillus niger]	gi|145230375	9 kDa									8	6
11			malate dehydrogenase (MDH_glycosomal_mitochondrial,NADB_Rossman) An07g02160 [Aspergill...	gi|145237310	36 kDa	*							5	8	4
12			Glyceraldehyde-3-phosphate dehydrogenase; GAPDH (NADB_Rossman, Gp_dh) [Aspergillus niger]	gi|2494634	36 kDa	*								7	
13			thioredoxin family protein (TRX_family) An01g02500 [Aspergillus niger]	gi|145228757	12 kDa								1	7	6
14			detected protein unk function 15_kDa An08g06620 [Aspergillus niger]	gi|145240261	15 kDa									7	7
15			Nascent polypeptide-associated complex subunit beta; Beta-NAC [Aspergillus niger]	gi|143355005	17 kDa	*							5	7	5
16			ATP synthase beta subunit, nucleotide-binding domain Pc21g10070 [Penicillium chrysogenum Wisc...	gi|211589753	55 kDa	*	9		9	8	0	3	5	6	3
17			calmodulin [Penicillium marneffei ATCC 18224]	gi|212538221	13 kDa	*			4	4	6	3		6	6
18			malate dehydrogenase (MDH_glycosomal_mitochondrial,NADB_Rossman) An15g00070 [Aspergillu...	gi|145250065	34 kDa	*							2	6	2
19			eukaryotic elongation factor 1 beta (EF1B) guanine nucleotide exchange domain An08g03490 [Asp...	gi|145239637	33 kDa									6	5
20			cell wall protein An14g02100 [Aspergillus niger]	gi|145248994	28 kDa								1	6	4
21			aspartic protease pepE [Aspergillus niger]	gi|145232965	43 kDa	*								5	
22			detected protein with Ran-binding domain 28_kDa (Ran8D) An18g02140 [Aspergillus niger]	gi|145254588	28 kDa	*								5	4
23			Catalase-peroxidase An01g01830 [Aspergillus niger]	gi|145228627	84 kDa	*								5	2
24			Nascent polypeptide-associated complex subunit alpha; NAC-alpha; Alpha-NAC [Neosartorya fische...	gi|143354992	22 kDa	*		1			1		3	5	4
25			detected protein unk function 21_kDa An09g00840 [Aspergillus niger]	gi|145241516	21 kDa									5	6
26			detected protein unk function 13_kDa An18g04120 [Aspergillus niger]	gi|145255128	13 kDa									5	4
27			conidial pigment biosynthesis oxidase (multicopper oxidases) An14g05370 [Aspergillus niger]	gi|145249642	62 kDa									5	2
28			alpha-glucosidase (Glyco_hydro_31) An01g10930 [Aspergillus niger]	gi|145230403	97 kDa	*								5	2
29			60S acidic ribosomal protein P2 An16g04930 [Aspergillus niger]	gi|145252576	11 kDa	*								5	5
30			similar to secreted dipeptidyl peptidase (Peptidase_S9,DAP2, Esterase_lipase superfamily) An12g...	gi|145246582	79 kDa								1	5	7
31			detected protein unk function 19_kDa An02g08300 [Aspergillus niger]	gi|145233179	19 kDa									5	5
32			detected protein unk function 55_kDa An07g05660 [Aspergillus niger]	gi|145237996	55 kDa									5	4
33			similar to cell wall galactomannoprotein An08g09420 [Aspergillus niger]	gi|145240813	20 kDa									5	2
34			Aldehyde dehydrogenase; ALDH [Aspergillus niger]	gi|1169291	54 kDa	*	1		0					4	1
35			detected lysine-rich protein 56_kDa An01g07870 [Aspergillus niger]	gi|145229807	56 kDa	*								4	5
36			Eukaryotic translation initiation factor 5A (eIF5A) An01g02900 [Aspergillus niger]	gi|145228831	18 kDa	*	2	1				2	3	4	5
37			molecular chaperone Hsp70 An07g09990 [Aspergillus niger]	gi|145238854	70 kDa	*	3						2	4	4
38			cyclophilin-like peptidyl prolyl cis-trans isomerase cypA [Aspergillus niger]	gi|145238516	19 kDa	*	1	1					1	4	4
39			enolase An18g06250 [Aspergillus niger]	gi|145255754	47 kDa		1						2	4	4
40			detected protein unk function 20_kDa An12g02680 [Aspergillus niger]	gi|145246196	20 kDa									4	6
41			transaldolase An07g03850 [Aspergillus niger]	gi|145237640	35 kDa	*							2	4	3
42			transketolase (TPP_TK) An08g06570 [Aspergillus niger]	gi|145240251	75 kDa	*							4	4	
43			Acyl CoA binding protein family An02g02960 [Aspergillus niger]	gi|145231936	16 kDa									4	3

Figure 5.10. An example view of a Scaffold™ results table compiled from LC/MS/MS analyses of eight fungal reference strains and the GAPLAB_RH isolate. Results have been sorted on Aspergillus niger NRRL 363 (column 8, category E) and the number of unique spectra detected. *A. niger* strains have many high quality matches to protein sequences in NCBI for *A. niger*, other species do not. Although *A. carbonarius* is closely related to *A. niger*, relatively few spectra from *A. carbonarius* were matched to peptides from *A. niger* proteins.

Sequencing of the ITS region of ribosomal DNA (rDNA). The products of the PCR amplification of genomic DNA with fungal ITS region primers were approximately 600 bp, as expected for Aspergillus species (Henry et al., 2000), based on agarose gel separation of the reaction products. Sequencing of the ITS region confirmed the identity of the *Aspergillus* isolates as *Aspergillus flavus*. The multiple sequence alignment of the *Aspergillus* reference strains and field isolates is presented in Figure 5.11. The sequence of the ITS region of isolate P98b matched U34555 *Fusarium verticillioides* with identities 462/464 (99%), using NCBI-BLAST2 to query EMBL Fungi, but also matched entries for other *Fusarium* species, such as *F. napiform*, *F. sacchari*, *F. bulbicola*, and *F. proliferatum*, with identical matching scores. Similarly, the sequence of the ITS region of isolate GAPLAB_RH matched EU861295 *Penicillium chrysogenum* with identities 394/396 (99%), but also matched entries of other *Penicillium* species, such as *P. citrinum*, *P. vinaceum*, and *P. oxalicum*, with identical matching scores.



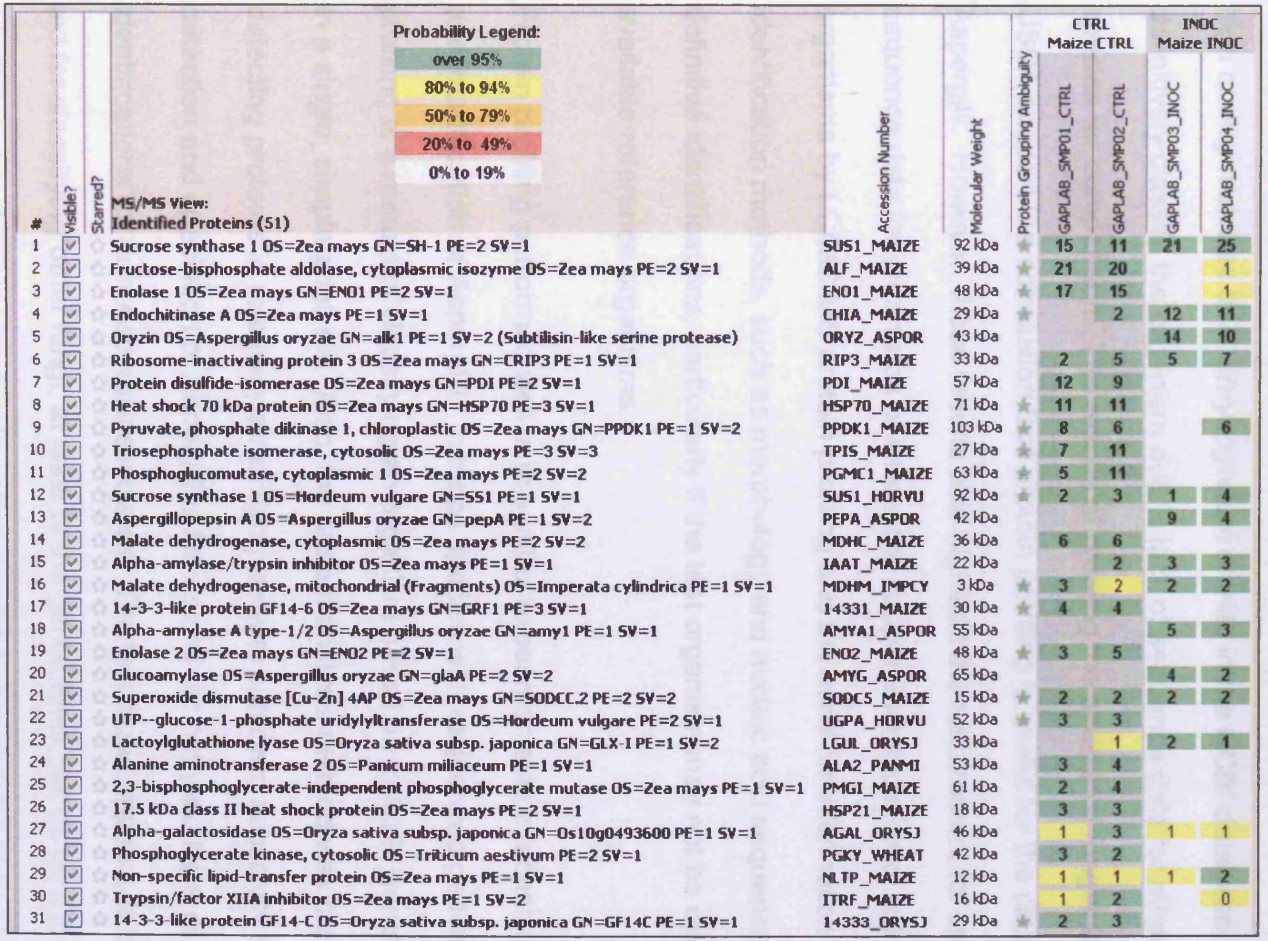
```
CLUSTAL 2.0.8 multiple sequence alignment

1957_ITS_consensus      AACCTCCCACCCGTGTTTACTGTACCTTAGTTGCTTCGGCGGGCCCGCCATTCATGGCCG 62
Clayton270_ITS          ----------------------------------TCGGCGGGCCCGCCATTCATGGCCG 25
3357_ITS4-ITS4.ab1      AACCTCCCACCCGTGTTTACTGTACCTTAGTTGCTTCGGCGGGCCCGCCATTCATGGCCG 119
Sandhills_ITS           AACCTCCCACCCGTGTTTACTGTACCTTAGTTGCTTCGGCGGGCCCGCCATTCATGGCCG 74
Plymouth_ITS4           AACCTCCCACCCGTGTTTACTGTACCTTAGTTGCTTCGGCGGGCCCGCCATTCATGGCCG 120
                                                  *************************

1957_ITS_consensus      CCGGGGGCTCTCAGCCCCGGGCCCGCGCCCGCCGGAGACACCACGAACTCTGTCTGATCT 122
Clayton270_ITS          CCGGGGGCTCTCAGCCCCGGGCCCGCGCCCGCCGGAGACACCACGAACTCTGTCTGATCT 85
3357_ITS4-ITS4.ab1      CCGGGGGCTCTCAGCCCCGGGCCCGCGCCCGCCGGAGACACCACGAACTCTGTCTGATCT 179
Sandhills_ITS           CCGGGGGCTCTCAGCCCCGGGCCCGCGCCCGCCGGAGACACCACGAACTCTGTCTGATCT 134
Plymouth_ITS4           CCGGGGGCTCTCAGCCCCGGGCCCGCGCCCGCCGGAGACACCACGAACTCTGTCTGATCT 180
                        ************************************************************

1957_ITS_consensus      AGTGAAGTCTGAGTTGATTGTATCGCAATCAGTTAAAACTTTCAACAATGGATCTCTTGG 182
Clayton270_ITS          AGTGAAGTCTGAGTTGATTGTATCGCAATCAGTTAAAACTTTCAACAATGGATCTCTTGG 145
3357_ITS4-ITS4.ab1      AGTGAAGTCTGAGTTGATTGTATCGCAATCAGTTAAAACTTTCAACAATGGATCTCTTGG 239
Sandhills_ITS           AGTGAAGTCTGAGTTGATTGTATCGCAATCAGTTAAAACTTTCAACAATGGATCTCTTGG 194
Plymouth_ITS4           AGTGAAGTCTGAGTTGATTGTATCGCAATCAGTTAAAACTTTCAACAATGGATCTCTTGG 240
                        ************************************************************

1957_ITS_consensus      TTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAGTGTGAATTGCAGAATTCCG 242
Clayton270_ITS          TTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAGTGTGAATTGCAGAATTCCG 205
3357_ITS4-ITS4.ab1      TTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAGTGTGAATTGCAGAATTCCG 299
Sandhills_ITS           TTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAGTGTGAATTGCAGAATTCCG 254
Plymouth_ITS4           TTCCGGCATCGATGAAGAACGCAGCGAAATGCGATAACTAGTGTGAATTGCAGAATTCCG 300
                        ************************************************************

1957_ITS_consensus      TGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCCTGT 302
Clayton270_ITS          TGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCCTGT 265
3357_ITS4-ITS4.ab1      TGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCCTGT 359
Sandhills_ITS           TGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCCTGT 314
Plymouth_ITS4           TGAATCATCGAGTCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCCTGT 360
                        ************************************************************

1957_ITS_consensus      CCGAGCGTCATTGCTGCCCATCAAGCACGGCTTGTGTGTTGGGTCGTCGTCCCCTCTCCG 362
Clayton270_ITS          CCGAGCGTCATTGCTGCCCATCAAGCACGGCTTGTGTGTTGGGTCGTCGTCCCCTCTCCG 325
3357_ITS4-ITS4.ab1      CCGAGCGTCATTGCTGCCCATCAAGCACGGCTTGTGTGTTGGGTCGTCGTCCCCTCTCCG 419
Sandhills_ITS           CCGAGCGTCATTGCTGCCCATCAAGCACGGCTTGTGTGTTGGGTCGTCGTCCCCTCTCCG 374
Plymouth_ITS4           CCGAGCGTCATTGCTGCCCATCAAGCACGGCTTGTGTGTTGGGTCGTCGTCCCCTCTCCG 420
                        ************************************************************

1957_ITS_consensus      GGGGGGACGGGCCCCAAAGGCAGCGGCGGCACCGCGTCCGATCCTCGAGCGTATGGGGCT 422
Clayton270_ITS          GGGGGGACGGGCCCCAAAGGCAGCGGCGGCACCGCGTCCGATCCTCGAGCGTATGGGGCT 385
3357_ITS4-ITS4.ab1      GGGGGGACGGGCCCCAAAGGCAGCGGCGGCACCGCGTCCGATCCTCGAGCGTATGGGGCT 479
Sandhills_ITS           GGGGGGACGGGCCCCAAAGGCAGCGGCGGCACCGCGTCCGATCCTCGAGCGTATGGGGCT 434
Plymouth_ITS4           GGGGGGACGGGCCCCAAAGGCAGCGGCGGCACCGCGTCCGATCCTCGAGCGTATGGGGCT 480
                        ************************************************************

1957_ITS_consensus      TTGTCACCCGCTCTGTAGGCCCGGCCGGCGCTTGCCGAACGCAAATCAATCTTTTTCCAG 482
Clayton270_ITS          TTGTCACCCGCTCTGTAGGCCCGGCCGGCGCTTGCCGAACGCAAATCAATCTNTTTCCAG 445
3357_ITS4-ITS4.ab1      TTGTCACCCGCTCTGTAGGCCCGGCCGGCGCTTGCCGAACGCAAATCAATCTTTTTCCAG 539
Sandhills_ITS           TTGTCACCCGCTCTGTAGGCCCGGCCGGCGCTTGCCGAACGCAAATCAATCTTTT-CCAG 493
Plymouth_ITS4           TTGTCACCCGCTCTGTAGGCCCGGCCGGCGCTTGCCGAACGCAAATCAATCTTTT-CCAG 539
                        ************************************************** ** ****

1957_ITS_consensus      GNTGACCTCGNATCAGG------- 499
Clayton270_ITS          GNTGACCTCG-------------- 455
3357_ITS4-ITS4.ab1      GT-GACCTCGATCAGGAGGANCNA 562
Sandhills_ITS           GTNGACCTCNGATCAGG------- 510
Plymouth_ITS4           GTGA--CCTCGATCAGGAGATCNA 561
                        *     *
```

Figure 5.11. Multiple sequence alignment of ITS regions from the *Aspergillus flavus* type strain NRRL 1957, the sequenced strain NRRL3357, and the *Aspergillus* field isolates.

Protein profiling of Apoplastic Wash Fluid from maize kernels. Profiling of proteins from AWF was conducted to determine if there was sufficient fungal protein in the AWF to detect and identify the fungal pathogen in host tissue. The analysis yielded 10,000 MS/MS spectra per sample from 50 µL of apoplastic fluid recovered from four maize kernels. When searched versus a fungal subset of UniProt/Swiss-Prot, a total of 51 proteins, from maize and *Aspergillus*) were identified in the AWF samples (Figure 5.12). Not only were *Aspergillus* proteins detected, but the types of peptides found were sufficient to classify the organism as *A. flavus*. For example, a subtilisin-like protease called oryzin was detected and yielded a tryptic peptide specific to *A. flavus/oryzae* (Table 5.4). In addition, the maize proteins detected are different between control and inoculated, as if in response to infection (note chitinase), and many of the fungal proteins detected are enzymes involved in digesting protein and polysaccharide from the host.

Genus species	tryptic peptides
Aspergillus clavatus	GQASTNYVYDTSAGAGTYAYVVDSGINVDHIEFQG**R**
Aspergillus flavus/oryzae	GQQSTDYIYDTSAGEGTYAYVVDSGVNVDHEEFEG**R**
Aspergillus fumigatus	GQASTDYIYDTSAGAGTYAYVVDSGINVNHVEFES**R**
Aspergillus nidulans	GEASTTYVYDTSAGEGTYAYVVDTGINADHEEFGG**R**
Aspergillus niger	GGSSTDYIYDDSAGEGTYAYVVDTGILATHNEFGG**R**
Aspergillus terreus	GQPSTDYIYDTNGGEGTYAYVVDTGINVDHEEFEG**R**

Table 5.4. Tryptic peptide of oryzin diagnostic for identification of Aspergilli. Cleavage sites at arginine residues are in **BOLD**.



Probability Legend:

over 95%	80% to 94%	50% to 79%	20% to 49%	0% to 19%

#	Visible?	Starred?	MS/MS View: Identified Proteins (51)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	CTRL Maize CTRL GAPLAB_SMP01_CTRL	CTRL Maize CTRL GAPLAB_SMP02_CTRL	INOC Maize INOC GAPLAB_SMP03_INOC	INOC Maize INOC GAPLAB_SMP04_INOC
1	☑	☆	Sucrose synthase 1 OS=Zea mays GN=SH-1 PE=2 SV=1	SUS1_MAIZE	92 kDa	★	15	11	21	25
2	☑	☆	Fructose-bisphosphate aldolase, cytoplasmic isozyme OS=Zea mays PE=2 SV=1	ALF_MAIZE	39 kDa	★	21	20		1
3	☑	☆	Enolase 1 OS=Zea mays GN=ENO1 PE=2 SV=1	ENO1_MAIZE	48 kDa	★	17	15		1
4	☑	☆	Endochitinase A OS=Zea mays PE=1 SV=1	CHIA_MAIZE	29 kDa	★		2	12	11
5	☑	☆	Oryzin OS=Aspergillus oryzae GN=alk1 PE=1 SV=2 (Subtilisin-like serine protease)	ORYZ_ASPOR	43 kDa				14	10
6	☑	☆	Ribosome-inactivating protein 3 OS=Zea mays GN=CRIP3 PE=1 SV=1	RIP3_MAIZE	33 kDa		2	5	5	7
7	☑	☆	Protein disulfide-isomerase OS=Zea mays GN=PDI PE=2 SV=1	PDI_MAIZE	57 kDa		12	9		
8	☑	☆	Heat shock 70 kDa protein OS=Zea mays GN=HSP70 PE=3 SV=1	HSP70_MAIZE	71 kDa	★	11	11		
9	☑	☆	Pyruvate, phosphate dikinase 1, chloroplastic OS=Zea mays GN=PPDK1 PE=1 SV=2	PPDK1_MAIZE	103 kDa	★	8	6		6
10	☑	☆	Triosephosphate isomerase, cytosolic OS=Zea mays PE=3 SV=3	TPIS_MAIZE	27 kDa	★	7	11		
11	☑	☆	Phosphoglucomutase, cytoplasmic 1 OS=Zea mays PE=2 SV=2	PGMC1_MAIZE	63 kDa	★	5	11		
12	☑	☆	Sucrose synthase 1 OS=Hordeum vulgare GN=SS1 PE=1 SV=1	SUS1_HORVU	92 kDa	★	2	3	1	4
13	☑	☆	Aspergillopepsin A OS=Aspergillus oryzae GN=pepA PE=1 SV=2	PEPA_ASPOR	42 kDa				9	4
14	☑	☆	Malate dehydrogenase, cytoplasmic OS=Zea mays PE=2 SV=2	MDHC_MAIZE	36 kDa		6	6		
15	☑	☆	Alpha-amylase/trypsin inhibitor OS=Zea mays PE=1 SV=1	IAAT_MAIZE	22 kDa			2	3	3
16	☑	☆	Malate dehydrogenase, mitochondrial (Fragments) OS=Imperata cylindrica PE=1 SV=1	MDHM_IMPCY	3 kDa	★	3	2	2	2
17	☑	☆	14-3-3-like protein GF14-6 OS=Zea mays GN=GRF1 PE=3 SV=1	14331_MAIZE	30 kDa	★	4	4		
18	☑	☆	Alpha-amylase A type-1/2 OS=Aspergillus oryzae GN=amy1 PE=1 SV=1	AMYA1_ASPOR	55 kDa				5	3
19	☑	☆	Enolase 2 OS=Zea mays GN=ENO2 PE=2 SV=1	ENO2_MAIZE	48 kDa	★	3	5		
20	☑	☆	Glucoamylase OS=Aspergillus oryzae GN=glaA PE=2 SV=2	AMYG_ASPOR	65 kDa				4	2
21	☑	☆	Superoxide dismutase [Cu-Zn] 4AP OS=Zea mays GN=SODCC.2 PE=2 SV=2	SODC5_MAIZE	15 kDa	★	2	2	2	2
22	☑	☆	UTP--glucose-1-phosphate uridylyltransferase OS=Hordeum vulgare PE=2 SV=1	UGPA_HORVU	52 kDa	★	3	3		
23	☑	☆	Lactoylglutathione lyase OS=Oryza sativa subsp. japonica GN=GLX-I PE=1 SV=2	LGUL_ORYSJ	33 kDa			1	2	1
24	☑	☆	Alanine aminotransferase 2 OS=Panicum miliaceum PE=1 SV=1	ALA2_PANMI	53 kDa		3	4		
25	☑	☆	2,3-bisphosphoglycerate-independent phosphoglycerate mutase OS=Zea mays PE=1 SV=1	PMGI_MAIZE	61 kDa		2	4		
26	☑	☆	17.5 kDa class II heat shock protein OS=Zea mays PE=2 SV=1	HSP21_MAIZE	18 kDa		3	3		
27	☑	☆	Alpha-galactosidase OS=Oryza sativa subsp. japonica GN=Os10g0493600 PE=1 SV=1	AGAL_ORYSJ	46 kDa		1	3	1	1
28	☑	☆	Phosphoglycerate kinase, cytosolic OS=Triticum aestivum PE=2 SV=1	PGKY_WHEAT	42 kDa		3	2		
29	☑	☆	Non-specific lipid-transfer protein OS=Zea mays PE=1 SV=1	NLTP_MAIZE	12 kDa		1	1	1	2
30	☑	☆	Trypsin/factor XIIA inhibitor OS=Zea mays PE=1 SV=2	ITRF_MAIZE	16 kDa		1	2		0
31	☑	☆	14-3-3-like protein GF14-C OS=Oryza sativa subsp. japonica GN=GF14C PE=1 SV=1	14333_ORYSJ	29 kDa	★	2	3		

Figure 5.12. An example view of a Scaffold™ results table compiled from LC/MS/MS analyses of two replicates each of AWF from control (uninfected) and infected (inoculated with *Aspergillus flavus*) maize kernels. Left two columns are control samples; right two columns infected. Note differences in proteins detected between control and infected maize, such as endochitinase (line 4) in infected maize.

Conclusions

The identities of field isolates tentatively identified as *A. flavus* were confirmed based on the large number of high quality matches to *A. flavus* entries in the reference database and the identifications were further supported by sequencing of the ITS regions of the isolates. Protein profiling supported the identification of the *Penicillium* isolate, GAPLAB_RH, as *Penicillium chrysogenum*, based on the large number of high quality matches to *P. chrysogenum* entries in the NCBI database, but does not definitively classify the organism due to lack of sequence data for other *Penicillium* species for comparison. The results of profiling the additional fungal strains from USDA support the conclusion that protein profiling is useful for the classification of Aspergilli. However, these results also highlight the requirement for specific sequence data for organisms of interest to support identification of unknown organisms by LC/MS/MS protein profiling and the need for information from other identification methods, such as morphology and nucleic acid sequencing, for definitive identifications, particularly if the test organism may not be closely related to available reference organisms.

Protein profiling performed well in this set of experiments as a technique for detection and identification of Aspergilli. In those species with sequenced genomes, protein profiling is a means of classifying an organism based on multiple genetic characters in a single, simple analysis. Also, the protein analysis method was capable of detecting proteins in relatively small samples with sufficient sensitivity to allow detection and speciation of *Aspergillus flavus* and did not yield false positive identifications for non-*Aspergillus* isolates. The technique appears to be useful as an aid for classifying other fungi, at least to the genus level, but so far it is unclear whether the technique can resolve other genera to the species level. In addition to

its value for detection and identification, protein profiling also provides additional biochemical information on the organism in the context of the test systems studied by identifying the predominant enzymes and structural proteins present under different experimental conditions.

Chapter 6: Discussion and Implications for Future Research

Discussion

Morphology and genetic sequence analysis are well established methods for classifying and identifying fungi. In this study, reference strains of different *Aspergillus* species, *Penicillium chrysogenum*, *Candida albicans*, and *Cryptococcus neoformans* were characterized using LC/MS and GC/MS biochemical profiling techniques in order to find specific small molecules, peptides or biochemical profiles that can be used in addition to established methods to detect and classify Aspergilli to the species level. Subsequently, analytical methods developed for characterizing the reference strains were applied, along with morphology and PCR, to characterize and identify several laboratory and field isolates.

GC/MS profiling did yield preliminary data that suggested that clustering and discriminant analysis of TMS-derivatized small molecules may be of some use for classifying or characterizing Aspergilli. However, it was clear that the technique was impractical for rapid, routine identification of these fungi, because classification by GC/MS profiling would require concurrent analysis of a large set of reference organisms to verify accurate groupings. Also, as with identification by morphology, identification by GC/MS profiling can be confounded by organisms with atypical phenotypes as was the case with classification of *A. terreus* NRRL 274 by GC/MS. This organism grouped correctly with other reference strains of *A. terreus* based on discriminant function analysis of the complete GC/MS profiling data set, but when a smaller data set based on polyols and trehalose was selected, the test sample of NRRL 274 appeared to have an unusual profile that segregated the strain from other

reference strains. GC/MS profiling was a useful technique for surveying the overall small molecule composition of the biomass and exploring biochemical differences between organisms, but was not suitable as a primary technique for identifying Aspergilli.

Targeted LC/MS/MS analyses of secondary metabolites, such as mycotoxins, was useful for confirming species identity in some cases, but mycotoxin production was not a reliable indicator because mycotoxins often require specialised media to optimize production and there were mycotoxigenic and non-mycotoxigenic strains within the same species. Small molecule profiling of the reference strains by LC/MS yielded a much richer data set than GC/MS profiling, but was less reproducible and quantitative. The resulting data was evaluated for qualitative comparisons. Small molecule analysis provided some supporting data for classifying Aspergilli, but these data were not suitable as primary indicators of genus or species. One peptide-like compound detected in spore washes appeared to be unique to *A. flavus* and *A. parasiticus*. Attempts to identify this compound by LC/MS/MS lead to development of method for analysis of protein the evaluation of protein profiling as a technique for identifying *Aspergillus* species.

Protein profiling by analysis of tryptic digests of fungal biomass produced high quality matches to a large number of fungal proteins and many of these proteins were correctly identified to the species of the reference strains. Although different organisms, tissue types, and growth conditions may produce different cross sections of proteins, many proteins yield tryptic peptides that are unique and can be used for classification to the species level. In effect, protein profiling matched the organisms to genus or species based on a large number of genetic characters without directly sequencing the DNA of the test organisms or limiting the evaluation to a few specific genetic characters as with PCR. However, because the technique relies on matching

spectra to existing sequences in a database, identification by protein profiling is limited to those fungi whose genomes have been sequenced or for which there are adequate nucleotide sequences available from ESTs. The utility of the technique will improve as sequencing is becoming much cheaper and more routine and more sequence data is becoming available in the public domain for a wider range of organisms. In fact, a recent review by Demirev and Fenselau (2008) describes the use of protein analysis by mass spectrometry for the rapid identification of *Bacilli.* Protein profiling should work well for organisms that have high quality sequence available and exhibit sufficient variability at the amino acid level. This was demonstrated in the case studies with the analysis of the *Fusarium* and *Penicillium* isolates. Protein profiling readily verified the genus of the *Fusarium* isolate, but identification of *Penicillium chrysogenum* was not possible until the sequence was published (van den Berg *et al.* 2008)for the organism and a FASTA file of the amino acid sequences of the proteins were available.

Identification by protein profiling is similar to identification by molecular techniques because differences in genetic sequences between organisms are the basis for classification. However, some of the target genes routinely used for classification by analysis of nucleic acid sequences can't be used for identification by amino acid sequence. Nucleotide sequences for a number of major proteins detected, such as actin, tubulin, Hsp70, GAPDH, enolase, and superoxide dismutase have been used for a number of phylogenetic studies (Gouy and Li 1989, Frisvad 1998, Samson *et al.* 2006, Yaguchi *et al.* 2007). Amino acid sequences of peptides from GAPDH, enolase, pyruvate decarboxylase, phosphoglycerate kinase, and alcohol dehydrogenase, along with a number of other proteins, provided high probability matches of fungi to correct genus and species. While these proteins did share peptide amino acid sequences across the species tested, these proteins were

present at sufficiently high levels that the peptide coverages were often greater than 20% and there were enough differences in the sequences that protein profiling of these proteins by LC/MS/MS could be used to correctly identify the test organisms to the species or at least section level. However, ATP synthases, ribosomal proteins, and chaperonins, as a rule, were relatively abundant and often had good sequence coverage, but the amino acid sequences of these proteins were relatively conserved. These sequences could be used to classify a fungus to the genus level, but not to the species level. Actin and tubulin were not major proteins detected in the total protein extract and the peptide coverages obtained for these proteins were often less than 20%. Also, protein profiling can only classify fungi based on transcribed sequences that code for proteins. The amplified sequences useful for taxonomy, such as actin and tubulin, contain non-coding regions that tend to show more variation than coding regions (Webster and Webster 2007). Similarly, the internal transcribed spacers (ITS) are transcribed, but do not code for proteins.

A disadvantage of classifying organisms by protein profiling compared to PCR is that protein profiling does not generate 'de novo sequence from peptide MS/MS spectra unless the software package(s) used for processing peptide MS/MS data employ special algorithms, such as SPIDER (Han *et al.* 2004) in PEAKS (Bioinformatics Solutions Inc.), that are capable of doing so. Commonly used software packages, such as Bioworks™ and Scaffold™ that were used in this study, rely on matching MS/MS spectra to theoretical spectra generated based on amino acid sequences in FASTA databases and do not generate high quality 'de novo' sequence based solely on interpretation of MS/MS spectra. Consequently, if only spectral matching is employed for data processing and a test organism has small genetic differences relative to sequenced strains, the resulting changes in amino acid sequences would generate mismatches between spectra for the test organism and the corresponding sequences in the reference database. For example, *A. parasiticus* is closely related

to *A. flavus* and *A. oryzae*, both sequenced, and many peptides and proteins match very well with those of the sequenced species, but in some cases the number of peptides identified and the confidences for protein and peptide identifications may be somewhat lower than for sequenced species. Likewise, *A. carbonarius* is closely related to *A. niger* and tends to group with *A. niger* based on GAPDH, pyruvate decarboxylase, and alcohol dehyrogenase peptides, but the overall matching was much more ambiguous than for test samples from sequenced species. *A. ochraceous* and *A. westerdijkiae* could not be classified by LC/MS/MS because nucleotide sequences specific to these organisms were not readily available and MS/MS spectra yielded lower confidence matches to peptides from other organisms in the database.

Other factors can influence the speed and quality of classification by protein profiling. Different matching algorithms (SEQUEST® in BioWorks™, X!Tandem in Scaffold™) generally yielded very similar protein matches, but specific peptide matches tended to vary between programs and the parameters used for processing. The reference FASTA database had a major influence on the quality of matches. A subset of NCBI sequences selected for fungi was the primary reference FASTA used for searches, but this large public database is not particularly well annotated and some errors in NCBI entries were noted during review. For example, in preparing multiple alignments, a large gap was noted in the NCBI entry for ATP synthase from *A. niger*. This required the manual addition of an ATP synthase sequence from the *A. niger* database of the U.S. Department of Energy Joint Genome Institute (JGI). In some cases, more effective searches may require searching versus FASTA databases specific to the target organism from alternate sources, such as the UniProt Knowledgebase, the Broad Insitute, EMBL-EBI, JGI, or The Institute for Genomic Research (TIGR), now part of the J. Craig Venter Institute (JCVI).

Although protein profiling proved to be a very powerful tool for identification of *Aspergillus*, identification of fungi is an iterative process and other factors were considered for final identification. Morphology was a primary factor in the preliminary identification of fungal isolates from the case study. There will be many fungi that do not present distinctive colours or structures, but when present, unique morphological characteristics can quickly narrow the list of possible matches. The source material is also a consideration in deciding which organism is a likely match. In the case study, both *Aspergillus flavus* and *Fusarium verticilioides* are commonly isolated from maize, so good preliminary identifications could be made based on which organisms were expected in the source material and the observed morphologies in culture. Protein profiling and ITS sequencing provided more definitive identification of the isolates.

Protein profiling is relatively fast (analysis can be accomplished in one day) and accurate for identification of *Aspergillus* and the simple processing method uses inexpensive reagents, but requires analysis on rather expensive mass spectrometers. PCR of the ITS region requires some expensive reagents and equipment, but inexpensive sequencing services are generally available in commercial and academic facilities. One of the major advantages of protein profiling over ITS sequencing, is that protein profiling provides a great deal of additional biochemical information on the organism by identifying the predominant enzymes and structural proteins present under different experimental conditions. This aspect of protein profiling will be exploited in ongoing studies with *Aspergillus* and other organisms.

Future Research

The CxP/DEA/HED protein extraction method developed for this study had been implemented as a standard method for protein analysis in the Genomic Sciences Laboratory at NC State University and has been effective in extracting proteins from a variety of other biological matrices, including bacteria, nematodes, daphnids, fish, plant leaves and fruit, mammalian and avian tissue, and for extracting proteins from agarose and acrylamide gels. The resulting protein samples can be used to prepare tryptic digests for LC/MS/MS analysis, as in this study, or with further processing, can be separated by SDS-PAGE. Additional method development is in progress to adapt the procedure for preparing samples suitable for two-dimensional gel electrophoresis.

In other studies that will be reported elsewhere, the extraction procedure has been shown to be compatible with protein analyses that use amine-reactive isobaric mass tags, such as iTRAQ® (Invitrogen) or Tandem Mass Tags (TMT®, Proteome Sciences) for relative quantitation. Isobaric tagging produces peptides with chemical labels such that identical peptides from differentially labelled samples have the same mass and chromatographic properties, but on MS/MS fragmentation yield different reporter ions (Thompson *et al.* 2003, Domon and Aebersold 2006, Aggarwal *et al.* 2006, Wiese *et al.* 2007). In the labelling process, the tryptic digest of each sample for comparison is labelled with its own unique tag, then all the samples are mixed and analysed in the same chromatographic run. The ratios of the reporter ions (a unique ion contributed by each labelled sample) are used to determine the relative abundances of peptides and the associated proteins in the samples. This will be particularly valuable for sample to sample comparisons of protein concentrations in time courses, induction/inhibition experiments, and studies of host-pathogen interactions.

Several follow up studies on *Aspergillus* are in progress to apply protein profiling to examine additional species in collaboration with the United States Department of Agriculture (USDA) National Center for Agricultural Utilization Research (NCAUR; Peoria, Illinois, USA) and the Centers for Disease Control and Prevention (CDC; Atlanta Georgia, USA). These studies may involve genomic sequencing, generating cDNA libraries, or cloning and sequencing specific genes because adequate sequence data may not be available for some of the organisms. Other studies are focused on further characterising host/pathogen interactions between *Aspergillus flavus* and maize using expression analysis (microarrays), targeted analyses, small molecule profiling and protein profiling, including the use of isobaric mass tags.

Conclusion

Biochemical profiling techniques based on GC/MS and LC/MS were successfully applied to survey the small molecule and protein composition of *Aspergillus*. While small molecule profiling generated a large volume of data and some interesting leads, small molecule profiling did not generate reliable biochemical patterns for definitively identifying Aspergilli, a major objective of the study. Protein profiling was a very effective technique for identifying *Aspergillus* species that have publicly available sequence data and the protein analysis methodology developed for this study has found wider applications for biochemical studies of *Aspergillus* and other organisms.

References

Aggarwal, K., L. H. Choe, and K. H. Lee. 2006. Shotgun proteomics using the iTRAQ isobaric tags. Brief Funct Genomic Proteomic 5:112-120.

Ahearn, D.G. 1978. Medically Important Yeasts. Annu. Rev. Microbiol. 32 (1): 59-68.

Ainsworth, G. C., Alfred S. Sussman, and F. K. Sparrow. 1965. The fungi; an advanced treatise. New York: Academic Press.

Alexopoulos, C.J. and C.W. Mims. 1979. Introductory mycology. New York: Wiley. xvii, 632 p.

Baldwin, M.A. 2004. Protein Identification by Mass Spectrometry: Issues to be Considered. Mol. Cell Proteomics 3 (1):1-9.

Bennett, J. W. and M. Klich. 2003. Mycotoxins. Clin. Microbiol. Rev. 16 (3):497-516.

Biron, D.G., C. Brun, T. Lefevre, C. Lebarbenchon, H.D. Loxdale, F. Chevenet, J-P. Brizard, and F. Thomas. 2006. The pitfalls of proteomics experiments without the correct use of bioinformatics tools. PROTEOMICS 6 (20):5577-5596.

Blackwell, Meredith, Rytas Vilgalys, Timothy Y. James, and John W. Taylor. 2008. Fungi. Eumycota: mushrooms, sac fungi, yeast, molds, rusts, etc.. Version 21 February 2008. http://tolweb.org/Fungi/2377/2008.02.21. In The Tree of Life Web Project, http://tolweb.org.

Botha, A.A. and J.L.F. Kock. 1993. Application of fatty acid profiles in the identification of yeasts. Int. J. Food Microbiol. 19 (1):39-51.

Christensen, M. 1981. A synoptic key and evaluation of species in the *Aspergillus flavus* group. Mycologia 73:1056.

Cline, E. 2005. Implications of changes to Article 59 of the International Code of Botanical Nomenclature enacted at the Vienna Conference 2005. Inoculum 56 (6):3-5.

Clutterbuck, A. J. 1994. Genetics of Fungi. In The Growing Fungus, edited by N. A. R. Gow and G. M. Gadd. London: Chapman & Hall.

Council for Agricultural Science and Technology (CAST). 2002. Mycotoxins : risks in plant, animal, and human systems. Ames, IA: Council for Agricultural Science and Technology.

da Silva, T. L. 1998. Cellular fatty acid profiles for the differentiation of *Penicillium* species. FEMS Microbiol. Lett. 164 (2):303.

Deacon, J.W. 1984. Introduction to modern mycology. Oxford ; Boston: Blackwell Scientific Publications. viii, 239 p.

Demirev, P.A. and Catherine Fenselau. 2008. Mass Spectrometry for Rapid Characterization of Microorganisms. Annu. Rev. Anal. Chem. 1 (1):71-93.

Denning, D. W. 1998. Invasive aspergillosis. Clin. Infect. Dis. 26 (4):781-803.

Domon, B., and R. Aebersold. 2006. Mass Spectrometry and Protein Analysis. Science 312:212-217.

Dunn, W. B., and D. I. Ellis. 2005. Metabolomics: Current analytical platforms and methodologies. TrAC Trends iAnal. Chem. 24:285-294.

El Menyawi, I. I., M. M. Wögerbauer, H. H. Sigmund, H. H. Burgmann, and W. W. Graninger. 2000. Identification of yeast species by fatty acid profiling as measured by gas-liquid chromatography. J. Chrom. B Biomed. Sci. Appl. 742 (1):13-24.

Fernie, A. R., R. N. Trethewey, A. J. Krotzky, and L. Willmitzer. 2004. Metabolite profiling: from diagnostics to systems biology. Nat. Rev. Mol. Cell. Biol. 5:763-769.

Fiehn, O. 2002. Metabolomics – the link between genotypes and phenotypes. Plant Mol. Biol. 48 (1):155-171.

Frisvad, J.C., P.D. Bridge, and D.K. Arora. 1998. Chemical fungal taxonomy. New York: Marcel Dekker.

Frisvad, J. C., P. Skouboe, and R. A. Samson. 2005. Taxonomic comparison of three different groups of aflatoxin producers and a new efficient producer of aflatoxin B1, sterigmatocystin and 3-O-methylsterigmatocystin, *Aspergillus rambellii* sp. nov. Syst. Appl. Microbiol. 28:442-453.

Galagan, J.E., S.E. Calvo, C. Cuomo, L-J. Ma, J.R. Wortman, S. Batzoglou, S-I. Lee, M. Basturkmen, C.C. Spevak, J. Clutterbuck, V. Kapitonov, J. Jurka, C. Scazzocchio, M. Farman, J. Butler, S. Purcell, S. Harris, G.H. Braus, O. Draht, S. Busch, C. D'Enfert, C. Bouchier, G.H. Goldman, D. Bell-Pedersen, S. Griffiths-Jones, J.H. Doonan, J. Yu, K. Vienken, A. Pain, M. Freitag, E.U. Selker, D.B. Archer, M.A. Penalva, B.R. Oakley, M. Momany, T. Tanaka, T. Kumagai, K. Asai, M. Machida, W.C. Nierman, D.W. Denning, M. Caddick, M. Hynes, M. Paoletti, R. Fischer, B. Miller, P. Dyer, M.S. Sachs, S.A. Osmani, and B.W. Birren. 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438 (7071):1105-1115.

Gams, W., M. Christensen, A.H.S. Onions, J.I. Pitt, and R.A. Samson. 1985. Infrageric taxa of *Aspergillus*. In Advances in *Aspergillus* Systematics, edited by R. A. Samson and J. I. Pitt. New York: Plenum Press.

Gams, W., and K.A. Seifert. 2008. Anamorphic Fungi. In Encyclopedia of Life Sciences (ELS). Chichester: John Wiley & Sons. http://www.els.net [doi: 10.1002/9780470015902.a0000351.pub2].

Geiser, D.M., C. Gueidan, J. Miadlikowska, F. Lutzoni, F. Kauff, V. Hofstetter, E. Fraker, C.L. Schoch, L. Tibell, W.A. Untereiner, and A. Aptroot. 2006. Eurotiomycetes: Eurotiomycetidae and Chaetothyriomycetidae. Mycologia 98 (6):1053-1064.

Geiser, D. M., M. A. Klich, J. C. Frisvad, S. W. Peterson, J. Varga, and R. A. Samson. 2007. The current status of species recognition and identification in Aspergillus. Stud Mycol 59 (1):1-10.

Geiser, David M. 2008. Sexual structures in *Aspergillus*: morphology, importance and genomics. Med. Mycol. 12:1 - 6. [Epub ahead of print]

Gil-Lamaignere, C., E. Roilides J. Hacker F. M. C. Müller. 2003. Molecular typing for fungi--a critical review of the possibilities and limitations of currently and future methods. Clin. Microbiol. Infect. 9 (3):172-185.

Glassbrook, N. and J. Ryals. 2001. A systematic approach to biochemical profiling. Curr. Opin. Plant Biol. 4 (3):186-190.

Gouy, M., and W. H. Li. 1989. Molecular phylogeny of the kingdoms Animalia, Plantae, and Fungi. Mol. Biol. Evol. 6 (2):109-122.

Guarro, J., J. Gene, and A.M. Stchigel. 1999. Developments in fungal taxonomy. Clin. Microbiol. Rev. 12 (3):454-500.

Gugnani, H.C. 2003. Ecology and taxonomy of pathogenic Aspergilli. Front. Biosci. 8:s346-57.

Halket, J.M., D. Waterman, A.M. Przyborowska, R.K.P. Patel, P.D. Fraser, and P.M. Bramley. 2005. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J. Exp. Bot. 56 (410):219-243.

Hamilton, A. J. and B. L. Gomez. 1998. The application of monoclonal antibodies to the diagnosis of disseminated mycoses. Rev. Iberoam. Micol. 15:118-124.

Han Y., B. Ma, and K. Zhang, 2004. SPIDER: software for protein identification from sequence tags with de novo sequencing error. J. Bioinform. Comput. Biol. (3): 697–716.

He, Z-M., Price, M., OBrian, G., Georgianna, D.R., and Payne, G.A. 2007. Improved protocols for functional analysis in the pathogenic fungus *Aspergillus flavus*. BMC Microbiology 7 (1):104.

Hedayati, M. T., A.C. Pasqualotto, P.A. Warn, P. Bowyer, and D.W. Denning. 2007. *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. Microbiology 153 (6):1677-1692.

Hennebert, G.L., and M. Vancanneyt. 1998. Proteins in Fungal Taxonomy. In Chemical fungal taxonomy, edited by J. C. Frisvad, P. D. Bridge and D. K. Arora. New York: Marcel Dekker.

Hibbett, D. S., M. Binder, J. F. Bischoff, M. Blackwell, P. F. Cannon, O. E. Eriksson, S. Huhndorf, T. James, P. M. Kirk, R. Lücking, H. Thorsten Lumbsch, F. Lutzoni, P. B. Matheny, D. J. McLaughlin, M. J. Powell, S. Redhead, C. L. Schoch, J. W. Spatafora, J. A. Stalpers, R. Vilgalys, M. C. Aime, A. Aptroot, R. Bauer, D. Begerow, G. L. Benny, L. A. Castlebury, P. W. Crous, Y.-C. Dai, W. Gams, D. M. Geiser, G. W. Griffith, C. Gueidan, D. L. Hawksworth, G. Hestmark, K. Hosaka, R. A. Humber, K. D. Hyde, J. E. Ironside, U. Kõljalg, C. P. Kurtzman, K.-H. Larsson, R. Lichtwardt, J. Longcore, J. Miadlikowska, A. Miller, J.-M. Moncalvo, S. Mozley-Standridge, F. Oberwinkler, E. Parmasto, V. Reeb, J. D. Rogers, C. Roux, L. Ryvarden, J. P. Sampaio, A. Schüßler, J. Sugiyama, R. G. Thorn, L. Tibell, W. A. Untereiner, C. Walker, Z. Wang, A. Weir, M. Weiss, M. M. White, K. Winka, Y.-J. Yao, and N. Zhang. 2007. A higher-level phylogenetic classification of the Fungi. Mycol. Res. 111:509-547.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. 66 (4):411-53.

Holstein, S.A. and R.J. Hohl. 2004. Isoprenoids: Remarkable diversity of form and function. Lipids 39 (4):293-309.

Hope, W. W., T. J. Walsh, and D. W. Denning. 2005. Laboratory diagnosis of invasive aspergillosis. Lancet Infect. Dis. 5 (10):609-622.

Humber, R. A. 2008. Evolution of entomopathogenicity in fungi. Journal of Invertebrate Pathology 98:262-266.

Jones, M.G. 2007. The first filamentous fungal genome sequences: *Aspergillus* leads the way for essential everyday resources or dusty museum specimens? Microbiology 153 (1):1-6.

Kell, D. B. 2004. Metabolomics and systems biology: making sense of the soup. Curr. Opin. Microbiol. 7:296.

Keller, A., A. I. Nesvizhskii, E. Kolker, and R. Aebersold. 2002. Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. Anal. Chem. 74:5383-5392.

Kirk, J. L., L. A. Beaudette, M. Hart, P. Moutoglis, J. N. Klironomos, H. Lee, and J. T. Trevors. 2004. Methods of studying soil microbial diversity. J. Microbiol. Methods 58 (2):169-88.

Klich, M.A. 2006. Identification of clinically relevant Aspergilli. Med. Mycol. 44 (0):127-131.

Klich, M. A., E. J. Mullaney, C. B. Daly, and J. W. Cary. 2000. Molecular and physiological aspects of aflatoxin and sterigmatocystin biosynthesis by *Aspergillus tamarii* and *A. ochraceoroseus*. Appl. Microbiol. Biotechnol. 53 (5):605-9.

Klont, R.R., M.A.S. H. Mennink-Kersten, and P.E. Verweij. 2004. Utility of *Aspergillus* Antigen Detection in Specimens Other than Serum Specimens. Clin. Infect. Dis. 39 (10):1467-1474.

Kobayashi, T., K. Abe, K. Asai, K. Gomi, P.R. Juvvadi, M. Kato, K. Kitamoto, M. Takeuchi, and M. Machida. 2007. Genomics of *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. 71 (3):646-670.

Kurtzman, C. P. 1994. Molecular taxonomy of the yeasts. Yeast 10 (13):1727-40.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

Latge, J-P. 1999. *Aspergillus fumigatus* and Aspergillosis. Clin. Microbiol. Rev. 12 (2):310-350.

Lohaus, G., K. Pennewiss, B. Sattelmacher, Hussmann M., and K.H. Muehling. 2001. Is the infiltration-centrifugation technique appropriate for the isolation of apoplastic fluid? A critical evaluation with different plant species. Physiologia Plant. 111:457–465.

Lutzoni, Francois, Frank Kauff, Cymon J. Cox, David McLaughlin, Gail Celio, Bryn Dentinger, Mahajabeen Padamsee, David Hibbett, Timothy Y. James, Elisabeth Baloch, Martin Grube, Valerie Reeb, Valerie Hofstetter, Conrad Schoch, A. Elizabeth Arnold, Jolanta Miadlikowska, Joseph Spatafora, Desiree Johnson, Sarah Hambleton, Michael Crockett, Robert Shoemaker, Gi-Ho Sung, Robert Lucking, Thorsten Lumbsch, Kerry

O'Donnell, Manfred Binder, Paul Diederich, Damien Ertz, Cecile Gueidan, Karen Hansen, Richard C. Harris, Kentaro Hosaka, Young-Woon Lim, Brandon Matheny, Hiromi Nishida, Don Pfister, Jack Rogers, Amy Rossman, Imke Schmitt, Harrie Sipman, Jeffrey Stone,

Lumbsch, H. T. 2000. Phylogeny of filamentous ascomycetes. Naturwissenschaften 87 (8):335-342.

Mackenzie, D.W.R. 1988. *Aspergillus* in man. In *Aspergillus* and Aspergillosis, edited by H. V. Bossche, D. W. R. Mackenzie and G. Cauwenbergh. New York: Plenum Press.

Matsuda, H., S. Kohno, S. Maesaki, H. Yamada, H. Koga, M. Tamura, H. Kuraishi, and J. Sugiyama. 1992. Application of ubiquinone systems and electrophoretic comparison of enzymes to identification of clinical isolates of *Aspergillus fumigatus* and several other species of *Aspergillus*. J. Clin. Microbiol. 30 (8):1999-2005.

McClenny, N. 2005. Laboratory detection and identification of *Aspergillus* species by microscopic observation and culture: the traditional approach. Med. Mycol. 43 Suppl 1:S125-128.

McDonald, B.A. 1997. The Population Genetics of Fungi: Tools and Techniques. Phytopathology 87 (4):448-453.

Mitterdorfer, G., H.K. Mayer, W. Kneifel, and H. Viernstein. 2002. Protein fingerprinting of *Saccharomyces* isolates with therapeutic relevance using one- and two-dimensional electrophoresis. Proteomics 2 (11):1532-1538.

Munkvold, Gary P. 2003. Epidemiology of Fusarium Diseases and their Mycotoxins in Maize Ears. Eur. J. Plant Pathol. 109 (7):705-713.

Murakami, H., K. Hayashi, and S. Ushijima. 1982. Useful key characters separating three *Aspergillus* taxa: *A. sojae*, *A. parasiticus*, and *A. toxicarius*. J. Gen. Appl. Microbiol. 28:55-60.

Murphy, P.A., S. Hendrich, C. Landgren, and C.M. Bryant. 2006. Food Mycotoxins: An Update. J. Food Sci. 71 (5):R51-R65.

Nemec, T. 1997. Sterols and fatty acids of different *Aspergillus* species. FEMS Microbiol. Lett. 149 (2):201.

Nesvizhskii, A. I., A. Keller, E. Kolker, and R. Aebersold. 2003. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. Anal. Chem. 75:4646-4658.

Nickerson, K.W., A.L. Atkin, and J.M. Hornby. 2006. Quorum Sensing in Dimorphic Fungi: Farnesol and Beyond. Appl. Environ. Microbiol. 72 (6):3805-3813.

Nolan, R. L., and J. K. Teller. 2006. Diethylamine extraction of proteins and peptides isolated with a mono-phasic solution of phenol and guanidine isothiocyanate. J. Biochem. Biophys. Methods 68:127-131.

Pasqualotto, A. C., and D. W. Denning. 2005. Diagnosis of Invasive Fungal Infections - Current Limitations of Classical and New Diagnostic Methods. Europ. Ocol. Rev. REV:1-11.

Paterson, R. R. M. 1998. Chemotaxonomy of Fungi by Unsaponifiable Lipids. In Chemical Fungal Taxonomy, edited by J. C. Frisvad, P. D. Bridge and D. K. Arora. New York: Marcel Dekker.

Peterson, S. W. 2000. Phylogenetic relationships in *Aspergillus* based upon rDNA sequence analysis. In Integration of Modern Taxonomic Methods for *Penicillium* and *Aspergillus* Classification, edited by R. A. Samson and J. I. Pitt. Singapore: Harwood Academic Publishers.

Peterson, S.W. 2008. Phylogenetic analysis of *Aspergillus* species using DNA sequences from four loci. Mycologia 100 (2):205-226.

Pickering, J.W., H.W. Sant, C.A.P. Bowles, W.L. Roberts, and G.L. Woods. 2005. Evaluation of a (1-3)-β-D-Glucan Assay for Diagnosis of Invasive Fungal Infections. J. Clin. Microbiol. 43 (12):5957-5962.

Pitt, J. I., and A. D. Hocking. 1985. Interfaces among Genera Related to *Aspergillus* and *Penicillium*. Mycologia 77:810-824.

Pitt, J. I. and R. A. Samson. 2007. Nomenclatural considerations in naming species of *Aspergillus* and its teleomorphs. Stud Mycol 59 (1):67-70.

Raper, K. B., and D. I. Fennell. 1965. The Genus Aspergillus. Williams & Wilkins, Baltimore, MD.

Ratna, Singh. 1973. The foot cell morphology of genus Aspergillus. Mycopathologia 49 (2):209-215.

Rath, P. M. 2001. Phenotypic and genotypic characterization of reference strains of the genus *Aspergillus*. Mycoses 44 (3-4):65-72.

Rinyu, E., J. Varga, and L. Ferenczy. 1995. Phenotypic and genotypic analysis of variability in *Aspergillus fumigatus*. J. Clin. Microbiol. 33 (10):2567-2575.

Rohringer, R., F. Ebrahim-Nesbat, and G. Wolf. 1983. Proteins in Intercellular Washing Fluids from Leaves of Barley (*Hordeum vulgare* L.). J. Exp. Bot. 34 (12):1589-1605.

Rokas, A., G. Payne, N.D. Fedorova, S.E. Baker, M. Machida, J. Yu, D.R. Georgianna, R.A. Dean, D. Bhatnagar, T.E. Cleveland, J.R. Wortman, R. Maiti, V. Joardar, P. Amedeo, D.W. Denning, and W.C. Nierman. 2007. What can comparative genomics tell us about species concepts in the genus *Aspergillus*? Stud. Mycol. 59 (1):11-17.

Samson, Robert A. 1992. Current taxonomic schemes of the genus *Aspergillus* and its teleomorphs. In *Aspergillus*: Biology and Industrial Applications, edited by J. W. Bennett and M. A. Klich. Boston: Butterworth-Heinemann.

Samson, R.A., S-B. Hong, and J.C. Frisvad. 2006. Old and new concepts of species differentiation in *Aspergillus*. Med. Mycol. 44 (0):133-148.

Scheidegger, K. A., and G.A. Payne. 2003. Unlocking the secrets behind secondary metabolism: A review of *Aspergillus flavus* from pathogenicity to functional genomics. Journal of toxicology. Toxin. Rev. 22 (2-3):423-459.

Seifert, K., and C. Lévesque. 2004. Phylogeny and Molecular Diagnosis of Mycotoxigenic Fungi. Eur. J. Plant Pathol. 110 (5):449-471.

Semighini, C.P., J.M. Hornby, R. Dumitru, K.W. Nickerson, and S.D. Harris. 2006. Farnesol-induced apoptosis in *Aspergillus nidulans* reveals a possible mechanism for antagonistic interactions between fungi. Mol. Microbiol. 59 (3):753-764.

Semighini, C.P., N. Murray, and S.D. Harris. 2008. Inhibition of *Fusarium graminearum* growth and development by farnesol. FEMS Microbiol. Lett. 279 (2):259-264.

Shea, J.M. and M. Del Poeta. 2006. Lipid signaling in pathogenic fungi. Curr. Opin. Microbiol. 9 (4):352-358.

Smith, Anthony D. 2000. Oxford dictionary of biochemistry and molecular biology. Oxford; New York: Oxford University Press.

Sorenson, W. G., H.W. Larsh, and S. Hamp. 1971. Acrylamide Gel Electrophoresis of Proteins from *Aspergillus* Species. Am. J. Bot. 58 (6):588.

Spatafora, J.W., G-H. Sung, D. Johnson, C. Hesse, B. O'Rourke, M. Serdani, R. Spotts, F. Lutzoni, V. Hofstetter, J. Miadlikowska, V. Reeb, C. Gueidan, E. Fraker, T. Lumbsch, R. Lucking, I. Schmitt, K. Hosaka, A. Aptroot, C. Roux, A.N. Miller, D.M. Geiser, J. Hafellner, G. Hestmark, A.E. Arnold, B. Budel, A. Rauhut, D. Hewitt, W.A. Untereiner, M.S. Cole, C. Scheidegger, M. Schultz, H. Sipman, and C.L. Schoch. 2006. A five-gene phylogeny of Pezizomycotina. Mycologia 98 (6):1018-1028.

Stchigel, A. M., and J. Guarro. 2007. A reassessment of cleistothecia as a taxonomic character. Mycol. Res. 111:1100-1115.

Sugiyama, J. 1998. Relatedness, phylogeny, and evolution of the fungi. Mycoscience 39 (4):487-511.

Sugiyama, J., R. Yahr, and R. Vilgalys. 2004. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. Am. J. Bot. 91 (10):1446-1480.

Thompson, A., J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, and C. Hamon. 2003. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. Anal. Chem. 75:1895-1904.

van den Berg, M.A., Albang, R., Albermann, K., Badger, J.H., Daran, J-M., Driessen, A.J.M., Garcia-Estrada, C., Fedorova, N.D., Harris, D.M., Heijne, W.H.M., Joardar, V., Kiel, J.A.K.W., Kovalchuk, A., Martin, J.F., Nierman, W.C., Nijland, J.G., Pronk, J.T., Roubos, J.A., van der Klei, I.J., van Peij, N.N.M.E., Veenhuis, M., von Dohren, H., Wagner, C., Wortman, J. and R.A.L. Bovenberg. 2008. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. Nat. Biotech. 26 (10):1161-1168.

Varga, J., K. Rigo, B. Toth, J. Teren, and Z. Kozakiewicz. 2003. Evolutionary Relationships among *Aspergillus* Species Producing Economically Important Mycotoxins. Food Technol. Biotechnol. 41 (1):29.

Varga, J., A. Juhasz, F. Kevei, and Z. Kozakiewicz. 2004. Molecular Diversity of Agriculturally Important *Aspergillus* Species. Eur. J. Plant Pathol. 110 (5/6):627.

Vélëz, H., N.J. Glassbrook, and M.E. Daub. 2007. Mannitol metabolism in the phytopathogenic fungus *Alternaria alternata*. Fungal Genet. Biol. 44 (4):258-268.

Vélëz, H., N.J. Glassbrook, and M.E. Daub. 2008. Mannitol biosynthesis is required for plant pathogenicity by *Alternaria alternata*. FEMS Microbiol. Lett. 285 (1):122-129.

Webster, J. and R. W. S. Weber. 2007. Introduction to Fungi. Cambridge University Press, Cambridge.

Wessels, J.G.H. 2005. Fungal physiology. In Encyclopedia of Life Sciences (ELS). Chichester: John Wiley & Sons. http://www.els.net [doi: 10.1038/npg.els.0004305]

Wiese, S., K. A. Reidegeld, H. E. Meyer, and B. Warscheid. 2007. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. PROTEOMICS 7:340-350.

Winter, P. C., G. I. Hickey, and H. L. Fletcher. 2002. Instant Notes in Genetics. 2nd ed. Oxford: BIOS Scientific Publishers.

Wogan, G.N. 1975. Mycotoxins. Annu. Rev. Pharmacol. 15:437-51.

Yaguchi, T. T., Y.Y. Horie, R.R. Tanaka, T. T. Matsuzawa, J.J. Ito, and K.K. Nishimura. 2007. Molecular phylogenetics of multiple genes on *Aspergillus* section Fumigati isolated from clinical specimens in Japan. Nihon Ishinkin Gakkai zasshi 48 (1):37-46.

Zelles, L. 1999. Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation of microbial communities in soil: a review. Biol. Fertil. Soils 29 (2):111-129.