Proteins and their Glycosylations as Diagnostic Biomarkers of Valley Fever

by

Natalie M. Mitchell

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2019 by the
Graduate Supervisory Committee:

Douglas F. Lake, Chair
Heather D. Bean
Thomas E. Grys
D. Mitch Magee

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Valley Fever (VF), is a potentially lethal fungal pneumonia caused by *Coccidioides spp.*, which is estimated to cause ~15-30% of all community-acquired pneumonias in the highly endemic Greater Phoenix and Tucson areas of Arizona. However, an accurate antigen-based diagnostic is still lacking. In order to identify protein and glycan antigen biomarkers of infection, I used a combination of genomics, proteomics and glycomics analyses to provide evidence of genus-specific proteins and glycosylations. The next goal was to determine if *Coccidioides*-specific glycans were present in biological samples from VF patients. Urine collected from 77 humans and 63 dogs were enriched for glycans and evaluated by mass spectrometry for *Coccidioides*-specific glycans and evaluated against a panel of normal donor urines, urines from patients infected with other fungi, and fungal cultures from closely related pneumonia-causing fungi. A combination of 6 glycan biomarkers was 100% sensitive and 100% specific in the diagnosis of human VF subjects, while only 3 glycan biomarkers were needed for 100% sensitivity and 100 specificity in the diagnosis of dog VF subject. Additionally, a blinded trial of 23 human urine samples was correctly able to classify urine samples with 93.3% sensitivity and 100% specificity. The results of this research provides evidence that *Coccidioides* genus-specific glycosylations have potential as antigens in diagnostic assays.

DEDICATION

To my late father, Joseph S. Mitchell Jr.

I wish you were here.

To my mother and personal cheerleader, Chrysoula Mitchell.

No amount of words would be enough to thank you.

To my husband and best friend, Thomas E. Pandelakis.

Thanks to your humor, love and support through this process, I'm still smiling.

ACKNOWLEDGMENTS

I would first like to thank my PhD advisor, and mentor, Dr. Douglas Lake. Your insistence upon excellence kept me challenged and your infectious optimism pushed me to achieve academic and professional goals I was scared to try. I am eternally grateful for the opportunity to work in your lab and to learn from you. Thank you.

I'm also very grateful to the members of my Committee. My gratitude to Dr. Thomas Grys, my second lab dad. Thank you for the guidance, funding and human clinical samples that made this project possible. Dr. D. Mitch Magee thank you for all the motivating conversations and Cocci expertise. You are so very generous with your time and supplies. Thanks also to Dr. Heather Bean for your keen insight and chromatography expertise.

Many thanks to all the people who made intellectual contributions to this work. Thanks to Andrew Sherrard for assistance with the Proteogenomics study. Much appreciation to Dr. Reed Cartwright and Adam Orr for creation of MS-Bidet R script which was instrumental to the glycan biomarker study. Thanks to Drs. Ben Madden, Surendra Dasari, Mike Holmes and Christine Charlesworth from the Proteomics Core, Mayo Clinic Rochester for helping me get my proteomics legs.

To my lab mates Alexa, Francisca, Kirsten, and Calvin thank you for the friendships and all you taught me. Amber, thank you for the sisterly love and friendship

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

x

CHAPTER I

INTRODUCTION

Etiology, epidemiology and pathophysiology of *Coccidioides* species

The fungal genus *Coccidioides* contains two species: *Coccidioides immitis* and *Coccidioides posadasii*, which are the etiological agents of coccidioidomycosis (Valley Fever; VF). *Coccidioides spp.* are dimorphic ascomycetes of the family Onygenaceae existing in a mycelial form in the soil (saprobic phase) and a yeast-like spherule form (parasitic phase) in infected mammalian hosts.[1] Listed as NIH NAIAD Class C organisms, the CDC recognizes *Coccidioides sp.* as an important re-emerging threat in the areas of their endemicity.[2] They are endemic to the desert regions of the American southwest, northern Mexico and parts of central and South America. Within these areas, arthroconidial spores from the soil are inhaled, having the potential to cause disease in a wide range of hosts. The life cycle of *Coccidioides spp.* is shown in Figure 1.

Tubular structure emerges

Endospore converts to mycelium

Mycelia

Arthroconidia form as alternate cells along the mycelia

Magnified

Arthroconidia 3-5µm

Conida may germinate and become mycelia

Endospore continues cycle

Day 5 (120h) Spherule releases endospores

Day 4 (96h) Endosphores form in the spherule

Day 3 (72h) Immature Spherule

Soil disturbed

Arthroconidia are inhaled

Day 1 (24h) Arthroconidia enlarge

Day 2 (48h) Free nuclear division

Figure 1: Life cycle of *Coccidioides spp.*. Mycelia form in the saprobic phase, which articulate and dissociate into arthroconidia. Arthroconidia are able to travel in the air, which is then inhaled by mammalian hosts. Within the hosts, the arthroconidia enlarge , become rounded into endospores, which multiply within a spherule. Spherules rupture in the parasitic phase, releasing new endospores which propagate more spherules in the host or can enter the saprobic phase if outside of the proper host microenvironment. Image taken from: R. G. Lewis, Eric; R. Bowers, Jolene; M. Barker, Bridget (2015): Dust Devil: The Life and Times of the Fungus That Causes Valley Fever, *PLOS Pathogens*.[3] Image used in accordance with the Creative Commons Attribution license.

Most mammals, including sea mammals, are susceptible to coccidioidomycosis and the disease has even been seen in some reptiles in the Phoenix Zoo. In domesticated animals, the majority of cases of coccidioidomycosis seen in veterinary practices are in canines, with an estimated annual incidence rate of 4%-6% among dogs in Pima and

Maricopa Counties, Arizona.[4] In humans, an overall prevalence rate of 42.6 per 100,000 people has been reported in endemic areas, with primary pulmonary Coccidioidomycosis being responsible for 15-29% of community acquired pneumonia cases in these areas.[5] Approximately 60% of infected individuals will have subclinical disease and do not seek medical attention.[6] The majority of the remaining 40% of infected individuals experience a symptomatic mild to moderate flu-like illness that may spontaneously resolve. However, around 10% of those infected will have severe pneumonia, and roughly 1% of patients will progress to life threatening disseminated infection, often in the bones, joints, skin or meninges, with or without primary pulmonary co-infection.[7]

Although considered an orphan disease, infecting <200,000 people per annum in the United States, VF is associated with high morbidity. In 2007, an Arizona Department of Health Services (ADHS) survey reported that symptomatic individuals experienced symptoms for a median of 120 days and of those employed at the time of the survey, 74% reported missing work for a median of 14 days.[8] In California alone, there were 25,217 hospitalizations in the years 2000-2011, with an average cost of $186 million per year.[9] In addition to hospitalization costs, antifungal treatment costs range from $2,000 to $20,000 per year per patient, and patients with severe disease often require physical therapy for profound fatigue and loss of lung capacity.[10] Typically, treatment of symptomatic patients requires long-term (3-6 months) antifungal treatment, but may

3

require years or even life-long treatment in cases of chronic progressive pneumonia, dissemination or in immunocompromised individuals.

Diagnosis of Coccidioidomycosis

Coccidioidomycosis is considered to have a variable presentation, as not all patients present with a typical clinical picture. Also, the symptoms of disease are similar to a multitude of other diseases; from cancer to tuberculosis.[11] Even when clinicians suspect VF, the currently available diagnostic tests are either invasive or have poor predictive power. Complicated diagnostic rubrics thus exist to aid clinicians in the diagnosis of this disease. A definitive diagnosis of Coccidiomycosis is currently reliant upon a positive sputum or bronchiolar lavage (BAL) culture, or a combination of symptoms with microscopically present spherules in a tissue biopsy or symptoms with more than one positive serology method.[12] Culture of the colorless fungus from biological specimens is considered the "gold standard" diagnostic technique, however, only around 6% of sputum 30-64% of BAL specimens[13] and 20-40% of disseminated cultures, including cerebral spinal fluid (CSF) return positive cultures,[14].[15,16] In addition to low positive culture rates, the fungus can take several weeks to grow and must be handled in a Biosafety Level 3 facility (BSL3), as it poses as a significant risk to laboratory personnel. Culturing, although a very specific indicator of disease, is not adequately sensitive and is too invasive to be used in initial screening.

Non-invasive screening of VF is currently performed using serology-based diagnostics, which includes immunodiffusion (ID) assays using tube precipitin (IDTP) and complement fixation antibodies (IDCF), complement fixation assay (CF) and enzyme immunoassays (EIAs; IgM &IgG).[13,17,18] These serological tests reveal either patient derived IgM or IgG antibodies made to *Coccidioides* antigens. However, IgM antibodies are detectable in only around 50% of patients by 1 week after symptom onset, and approximately 90% by 3 weeks after symptom onset. Therefore, around 10% of patients never mount detectable IgM levels.[19] IgG antibodies are generally detectable by 4-6 weeks post symptom onset, and around 85-90% of patients have detectable IgG by 3 months, but this is a long time to wait for a diagnosis and by that time, the patient may have already resolved the infection. With high false negative rates, serological assays are not good stand-alone diagnostics. False positives are also seen in patients infected with closely related fungi.[20] Multiple rounds of serological testing using multiple methodologies are thus performed and results are used as a method to monitor disease progression/ resolution.[10]

Additionally, immunocompromised patients–the patients most likely to have severe coccidioidomycoses–have different serological profiles than immunocompetent hosts.[21-23] In a study of 298 immunocompetent and 62 immunosuppressed persons with symptomatic infection, the immunosuppressed patients had lower rates of seropositivity for every type of serological test during the first year after onset of symptoms.[22] Patients

with disseminated versus pulmonary infections; CSF etc. have different diagnostic and serological pictures as well.[15,24] Serological testing is thus not a standalone testing strategy.

Although not routinely used, a few antigen-based diagnostic tests for VF also exist. One such test is an EIA measuring galactomannan, a cell wall polysaccharide released during spherule rupture. Galactomannan is a (1-4)-linked beta-D-mannopyranose backbone linked to alpha-D-galactose branches. Within severely ill patients, approximately 70.8% of urines were positive, however 58.3% of the positive urines also cross reacted with a *Histoplasma* galactomannan EIA.[25] This galactomannan EIA was also evaluated with Coccidioidal dog urines and sera, but sensitivity was extremely poor (<20%).[26] An alternative Coccidioidal antigen detection assay that has been evaluated is an (1→3)-β-D-glucan assay (beta glucan; BG). Fungal BG is another important cell wall polysaccharide and is comprised of polymeric D-glucose, which has been shown to be involved in immune modulation during infection.[27,28] BG has been found in VF patient serum, urine and CSF, predominantly in either severely ill or disseminated infections.[29,30] In a limited study of 12 severely ill patients, BG was seen as sensitive (~92%), but not specific, as it was highly reactive with 91% of *Aspergillus* positive sera.[14,31] Due to the disease-state variability of performance of these assays and high cross-reactivity with other fungi, the clinical utility of current antigen testing platforms is limited.

Barring rare culture positive samples, it is clear that current VF diagnostic methodologies lack sufficient diagnostic accuracy to be considered stand-alone tests. Additionally, an antigen detection assay that is able to accurately diagnose acute infections is greatly needed.

Eukaryotic N- and O-glycosylation of proteins

Monosaccharides are the basic subunit of all carbohydrates. With a few exceptions, they generally have the formula $C_nH_{2n}O_n$. All monosaccharides are reducing sugars, containing a free aldehyde group (aldoses) or the ability to tautomerize into a form with a free aldehyde group (ketones). Monosaccharides can exist either in ring structures or in an open chain (rare). Their reducing end is free to other sugars via glycosidic linkages, forming chiral centers. Linkages can be either of alpha or beta orientation based on the oxygen in relation to the anomeric carbon. Although the resultant chemical formulas are the same, the anomers have different structures which can confer differences in biological properties. It has been calculated that three hexoses could produce anywhere from 1,056 – 27,648 unique trisaccharides, depending on factors considered.[32] The more monomeric units in a polysaccharide, the greater the complexity. For examples, a polysaccharide with six hexose sugars could have over 1 trillion possible combinations.

Due to the variety of monosaccharides and the unfathomable number of possible combinations thereof, sugars constitute one of the most structurally diverse biological substrates in existence. As a comparison to protein diversity, which is based on a uniform linkage (peptide bonds) of 20 naturally occurring amino acids in linear chains folding into 3D structures, glycan diversity is generated by over 20 monosaccharides that can be uniformly linked or branch at any individual monosaccharide into large 3D structures. Similar to the large proteins like titin, which is around 33,000 amino acids in length,[33] polysaccharides like cellulose can also be large, having been observed to contain over 15,000 glucose units in a single molecule in plants like cotton.[34] Further complicating matters, polysaccharides, also called glycans, can be attached to virtually all types biological molecules, from proteins to lipids and nucleic acids.

It is estimated that between 50% and 70% of eukaryotic proteins are glycosylated.[35] Glycosylation is a post-translational addition of sugars to proteins which is mediated by specific enzymes.[36] Glycosylation is often confused with glycation, the non-enzymatic addition of sugars to proteins by covalent bonding to free reducing ends, such as what happens to hemoglobin in the bloodstream of diabetic patients.[37] There are 5 types of glycosylation based on the sugar-peptide bond produced; N-, O- and C-linked glycosylation, glypation and phosphoglycosylation.[36,38] However, N-linked and O-linked glycosylations are the most well studied. Glycosylation is not uniform at a particular site

8

on a protein. At any given glycan-protein binding site, there can be over a hundred different glycan structures.[35] Glycan diversity is thus prodigious.

In N-glycosylation, binding of a preassembled 14-mer oligosaccharide to the protein occurs within the endoplasmic reticulum (ER) to a motif-dependent asparagine (Asn-X-Ser/Thr), with subsequent enzymatic glycan trimming and/or addition modifications occurring within the Golgi apparatus. On the other hand, O-glycosylation can occur in the ER, Golgi, cytosol and nucleus. It can even happen on proteins which were previously N-glycosylated in the ER. Unlike N-glycosylation, O-glycosylation starts with the attachment of a single N-acetylglucosamine by dolichol-phosphomannose-dependent mannosyltransferases, and each additional monosaccharide is added sequentially by a concert of other enzymes. Additionally, O-glycosylation does not have an amino acid consensus sequence, making glycosylation prediction difficult.

Each of the steps in N- and O-glycosylation are performed by specialized carbohydrate active enzymes (CAZymes). CAZymes include enzymes for the synthesis, degradation and modification of carbohydrates in all living organisms.[39] There are 5 CAZyme classes based on function: 1) Glycoside hydrolases (GH), 2) Glycosyltransferases (GT), 3) Polysaccharide lyases (PL), 4) Carbohydrate esterases (CE), and 5) Auxiliary activities (AA). Within each CAZyme class, there are multiple families. CAZyme families are structurally-related enzymes or functional domains. As sequence identity of non-homologous species is low, functional prediction based on

sequence alignment is difficult.[40] The current best classification system is CAZy (www.cazy.org).[41] Within the CAZy database, there are currently over 300 CAZyme families and 100,000 non-redundant enzymes.

Fungal cell wall composition

In general, the fungal cell wall is composed of different types of linear and branched polysaccharides and proteins, often with various posttranslational glycans and lipids (Figure 2). However, considerable differences in the polysaccharide composition of fungal cell walls have been observed in different fungi. For example, the cell wall of *C. albicans* is mainly composed of β-1,3-glucan branched with β-1,6-glucan, N- and O-mannosylated proteins and some chitin,[42] whereas the cell wall of *A. fumigatus* lacks β-1,6-glucan but is rich in α-1,3-glucan, β-1,3/1,4-glucan and proteins modified with N- and O-linked galactomannans.[43] *C. neoformans* is composed of β-1,3-glucan, β-1,6-glucan and α-1,3-glucan and the cell wall proteins are modified by asparagine-linked xylomannans.[39] The dimorphic fungus, *Histoplasma capsulatum,* possesses a yeast-phase cell wall composed of chitin, soluble galactomannan, β-1,3-glucan and α-1,3-glucan.[44]
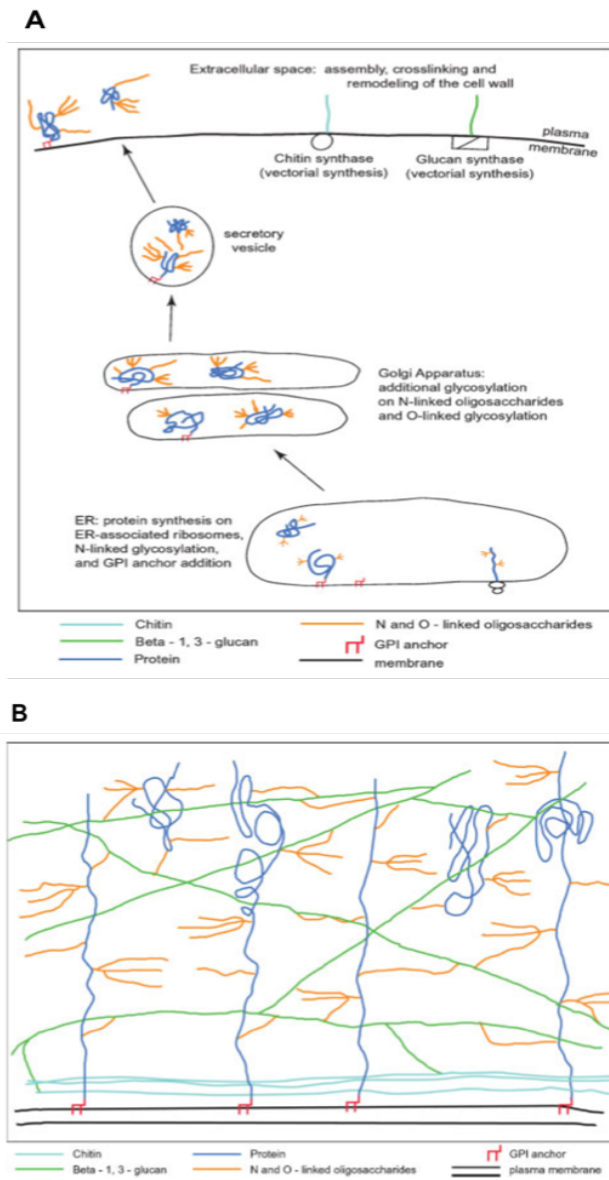
Figure 2: Protein glycosylation and the fungal cell wall. Proteins are produced and N-glycosylated in the ER prior to modification and O-glycosylation in the Golgi (A). Glycosylated proteins are either secreted or anchored to the fungal cell wall by Glycosylphosphatidylinositol (GPI) anchors (B). Modified from Bowman, S. M. and Free, S. J. (2006), The structure and synthesis of the fungal cell wall. *Bioessays*, 28: 799-808. doi:10.1002/bies.20441, with permission from John Wiley and Sons publishing.

11

Relatively little is known about the polysaccharide composition of the cell wall of *Coccidioides* spp., although it is known to contain chitin, galactomannan, β-1,3-glucan and 3-O-methylmannose.[45] It is also known that the composition of mycelial phase and spherule phase cell walls differ. Mycelial phase extracts have been recorded as containing 9.1-14.6% lipid,[46] 3-4% nitrogen in the form of protein,[47] and 35-70% polysaccharide,[47] as a percentage of dry weight. In the spherule phase, the lipid content has been observed as 23% of the weight of whole spherules.[46] *Coccidioides* spp. spherules also contain an outer cell wall (SOW), rich in lipids and glycolipids.[48] The neutral carbohydrate, nitrogen, and lipid content of the SOW have been reported as 16.7, 64.7, and 10.5%, respectively.[49]

Glycans as biomarkers in infectious diseases

Glycans are abundant in most biological fluids, including serum, plasma, saliva, urine and tears.[50] The application of glycomics to the diagnosis of disease is an active area of research in a variety of clinical diseases, from cancer to diabetes to allergy and rheumatoid arthritis.[51-55] Yet, to date, there are few studies utilizing microbial glycans for the diagnosis of disease. One microbe that has been evaluated for diagnostic glycans is *Mycobacterium tuberculosis* (TB). Lipoarabinomannan (LAM) is TB cell wall glycolipid that is shed from the lungs into the bloodstream and urine, and is seen as proportional to bacterial burden.[56] Lateral flow and ELISA test kits are now commercially available for

12

the diagnosis of TB using urinary LAM. Glycan biomarkers thus hold considerable opportunities for disease diagnosis.

As previously mentioned, the abundant fungal cell wall polysaccharides BG, chitin, and galactomannan are currently used as biomarkers of fungal disease. However, these sugars are composed of simple repeats produced by a limited number of CAZymes. They are therefore compositionally similar in all fungi. Fungal diagnostics that utilize BG and galactomannan in diagnostic platforms suffer from poor specificity, as they are not specific for a particular fungal genus. *Identification of glycans from N- or O-linked glycosylations, which involve the cooperation of a multitude of CAZymes, may provide more fungal-specific glycans.*

Lectin, Porous Graphitic Carbon and C18 binding properties

Lectins are carbohydrate binding proteins found ubiquitously throughout all kingdoms of life. They have specificity to particular sugars or motifs thereof. Although most lectins lack enzymatic activity, they serve a variety of functions. From acting as insecticides in plants to immune modulators in animals, their primary role is to serve in cell-cell contact. Lectin interactions with target carbohydrates is relatively weak, yet this interaction has been compared to Velcro; combined weak forces form strong complexes.[57] The vast number of lectins and their unique binding to glycoproteins and glycolipids make them attractive targets for use in medical research.

13

Lectins may be used to selectively enrich specific sugars or sugar motifs in complex biological samples. By covalently binding lectins to solid supports, like agarose beads, immunoprecipitations can be performed. Lectins are also employed in microarrays for the rapid identification of glycosylations and for glycoprotein profiling. In fact, lectin microarrays have been successfully used to identify cancer biomarkers in several cancer types.[58] One microarray containing 43 lectins was able to differentiate adenocarcinoma tissues from normal tissues.[59] In this microarray, wisteria floribunda agglutinin (WFA) was found to clearly differentiate cancerous from normal epithelia. WFA preferentially binds the disaccharide LacdiNAc motif (β-d-GalNAc-[1→4]-d-GlcNAc), which is overexpressed in certain cancers.[60]

On the other hand, porous graphitic carbon (PGC) and C18 are less selective than lectins and enrich according to chemical properties of the analyte, rather than binding motif.[61] Both PGC and C18 are frequently used in reversed phase high performance liquid chromatography (HPLC) or solid phase extraction (SPE) columns. However, their chemical composition and binding properties differ. The bonded phase of C18 columns is octadecylsilane, a chain of 18 carbon atoms to silica, whereas for PGC columns, it is graphitized carbon. PGC is made by carbonization of phenol-formaldehyde in porous silica at 1000°C, then the silica is removed prior to graphitization at 2000-2800°C.[62] PGC does not have the structure of graphite, which is arranged in well-organized three dimensional crystalline sheets. Instead, extensive layers of hexagonal carbon sheets are

stacked into a flat, homogenous surface. PGC and C18 binding are both based on

hydrophobic interactions, however PGC retention also increases with the number of

hydroxy groups. Unlike C18, PGC is thereby able to retain highly polar, hydrophilic

compounds, which is normally challenging in chromatography.[63] Additionally, owing to

the planar structure, PGC is able to separate isomeric compounds. PGC is thereby better

able to retain small hydrophilic molecules, while large hydrophobic molecules are easily

captured by C18.[63] They are thereby distinct yet complementary stationary phases used in

liquid chromatography fractionation.

Preliminary antigen enrichment experiments and rationale

In 2015, we reported that two lectins, succinylated wheat germ agglutinin

(sWGA) and *Griffonia simplifica* lectin II (GSLII), were able to specifically bind to

Coccidioides spp. spherules and not bind to surrounding human lung tissues in

immunohistochemistry (IHC) stains (Figure 3).[64] Using this preliminary evidence, it was

hypothesized that sWGA or GSLII lectins could be applied to enrich Coccidioidal

glycoproteins from *Coccidioides spp.* culture lysates. This hypothesis was tested against a

mycelial phase toluene autolysate from *Coccidoides posadasii* strain Silveira, termed

"coccidioidin" (CDN), prepared as described in Mitchell, NM, *et. al.* 2017. Using a

glycoprotein estimation kit (Thermo Scientific), CDN was estimated to contain between

15%-30% glycoprotein. Multiple attempts at enriching 5μg of CDN in phosphate

buffered saline (PBS), a concentration determined to be the upper limit of a physiologically relevant antigen amount which could be found circulating in human plasma, were performed. As exemplified in Figures 3A and 3B, WGA magnetic beads (Magnezoom beads, BioWorld Inc.) were able to enrich 5 μg of positive control glycoprotein, GlycNAc-BSA, but were unable to enrich 5μg of CDN. In-solution trypsin digestion and tandem mass spectrometry of elution fractions resulted in few peptide identifications, all of which were to peptides homologous to human peptides or to other medically relevant fungi (Figure 4C). Multiple attempts at lectin enrichments of CDN and enrichment with CDN affinity-purified goat anti-CDN IgG antibody-bound magnetic beads and various lectin beads, provided similarly disappointing results.
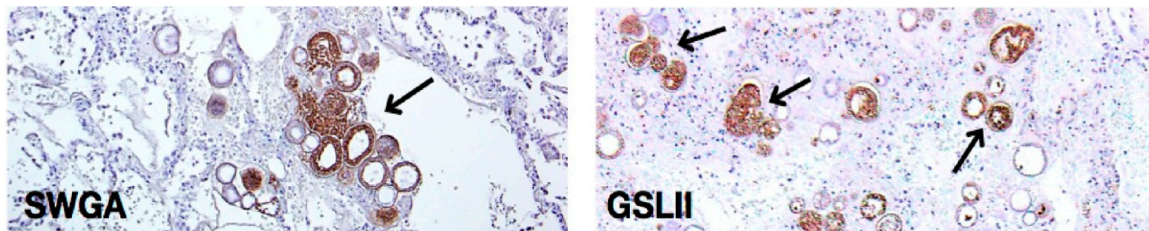
Figure 3: Succinylated wheat germ agglutinin (sWGA) and Griffonia simplificolia lectin II (GSLII) lectins selectively bind to *Coccidioides spp.* spherules within infected human lung tissues by immunohistochemistry staining. Reprinted with permission from Grys TE, et. al., Total and Lectin-Binding Proteome of Spherulin from *Coccidioides posadasii, J. Proteome Res*. 2016, 15, 10, 3463-3472. Copyright 2016, American Chemical Society.

38kDa WGA heterodimer ↓

A - PBS spiked with 5ug GlycNAc-BSA

B - PBS spiked with 5ug CDN

ladder | Flow through | Wash 1 | Wash 2 | Wash 3 | Wash 4 | Wash 5 | Elution 1 | Elution 2 | Elution 3

ladder | Flow through | Wash 1 | Wash 2 | Wash 3 | Wash 4 | Wash 5 | Elution 1 | Elution 2 | Elution 3

C

Protein	sWGA Agarose Beads				WGA Magnetic Beads			
	Normal donor plasma	Patient 1	Patient 2	5µg CDN in Normal donor plasma	Normal donor plasma	Patient 1	Patient 2	5µg CDN in Normal donor plasma
Elongation factor 1-alpha	2	0	1	3	1	1	1	1
Chperone protein dnaK	0	0	0	1	0	0	0	0
Elongation factor Tu	1	1	0	1	3	3	3	3
Histone H2B	1	0	0	1	1	1	1	1
ATP synthase alpha	0	0	0	2	2	1	0	2
ATP synthase beta	0	0	2	1	1	2	2	2
Adenosylhomocysteinase	3	1	0	0	0	3	3	5
Histone H2A	1	1	1	0	1	1	1	1

Figure 4. Lectin-bound agarose and magnetic beads are unable to enrich *Coccidioides spp.* glycoproteins from low concentrations of coccidioidin (CDN) or patient plasma. WGA magnetic beads (Magnezoom beads, BioWorld Inc.) were able to enrich 5 μg of positive control glycoprotein, GlycNAc-BSA (**A**), but were unable to enrich 5μg of CDN (**B**). Mass spectrometry of elution fractions from sWGA agarose beads and WGA magnetic beads resulted in few peptide identifications, all of which were to peptides homologous to human peptides or to other medically relevant fungi (**C**).

An additional issue noted during mass spectrometry was that, on average, only around 10% of CDN peptide spectra matched a peptide in the reference proteome for *C. posadasii* strain Silveira, the strain used to make CDN. As missed spectral hits can reduce the number of protein biomarkers identified, it is important to resolve the cause. It was first hypothesized that the post translational modifications, glycosylation, could add mass to the peptides, increasing their mass to charge (m/z) values, creating spectra that do not match the proteome database. However, following N-glycan deglycosylation of $20\mu g$ of CDN with PNGase F (New England Biolabs), used as per manufacturer instructions, no new proteins were identified and few additional peptides were identified by mass spectrometry (Figure 5). Glycosylation was thereby seen as an unlikely cause of missing spectra. The subsequent hypothesis was that missing spectra were due to spectra not matching the reference proteome database due to poor genome sequencing quality of *C. posadasii* strain Silveira. Sequencing or contig alignment errors in genomes can have the downstream effect of skipping protein-coding genes. Therefore, if the lectins were enriching proteins that were not in the reference proteome, one would miss those potential biomarkers of infection. Chapter II of this dissertation is titled "*Proteogenomic re-annotation of Coccidioides posadasii strain Silveira,*" and provides evidence that a large number of protein-coding genes are indeed missing from the current genome annotation of strain Silveira.

18
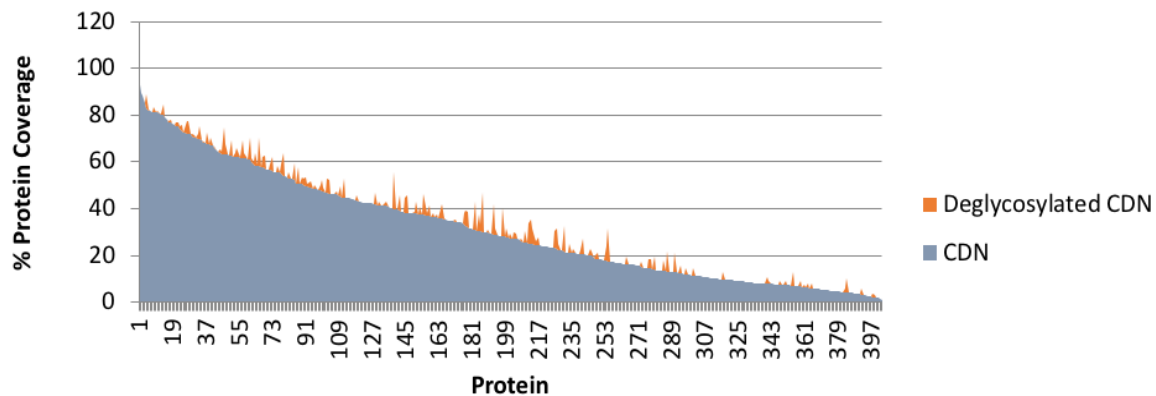
Figure 5. N-glycan deglycosylation of coccidioidin (CDN) does not provide more protein identifications or significantly improve percentage of protein coverage of *Coccidioides* proteins. CDN or CDN N-glcan deglycosylated with PNGase F overnight at 37°C was subjected to in-gel trypsinized and tandem mass spectrometry for comparison of protein identifications.

As lectin enrichment of glycoproteins were unfruitful, I was curious to determine the major protein antigens produced *in vivo* by *Coccidioides spp*. in lung tissues. As Coccidioidal cell walls are constantly being restructured and rupture to release endospores during infection, it is likely that the most abundant proteins produced *in vivo,* would be the most abundant proteins released into patient blood and urine. Human, mouse, dog and llama lung tissues infected with *Coccidioides spp*. were cut by laser capture microdissection (LCM) prior to mass spectrometry and estimation of protein abundance. Additionally, by comparing *in vivo*-grown with *in vitro-* grown *Coccidioides spp*., evidence is provided for which culture conditions and media produce the most *in-vivo*-like protein expression. This is written in Chapter III, titled *"LCM assisted protein*

biomarker discovery from Coccidioides-infected lung tissues and their in vitro

production."

The failure to enrich glycoproteins with lectins was initially puzzling, but

previous reports indicate that the Coccidioidal cell walls contains only 3-4% protein, 20-

25% lipid/glycolipid and over 50-60% carbohydrate, of which approximately 45% is

chitin/ chitosan.[47,65] Fungal polysaccharides are therefore abundant, and it is known that

Coccidioidal glycans like BG and galactomannan are released into patient blood and

urine. It is therefore possible that other Coccidioidal glycans are also released. However,

the utility of these glycans would be determined by whether they are specific to

Coccidioides spp. and if they are released in sufficient quantities to be detectable by

current detection methods. Preliminary evidence is mounting that genus or even species-

specific glycosylations in fungi exist. Interspecies variations with novel glycan

branching[66]and N-glycan synthesis due to absence or presence of certain

mannosyltransferases has been shown in filamentous fungi.[67] Coccidioides genus-specific

CAZymes and resultant glycans are evaluated in Chapter IV, titled "*Carbo-loading in*

fungal lung pathogens: a quantitative analysis of CAZyme abundance and resulting

glycan populations."

SWGA and GSLII lectins are known as chitin-binding lectins, and preferentially

bind to polymers of N-acetylglucosamine. It was therefore hypothesized that these lectins

were being overwhelmed by fungal cell wall chitin in the CDN lysate and chitin was

occupying lectin binding sites, preventing them from binding to other Coccidioidal glycoproteins. To test this hypothesis, an established alkaline lysis protocol was used to remove chitin and chitosan from CDN.[68] A lectin ELISA showed binding preference to chitin-containing fractions of CDN (Figure 6), with very little binding of chitin-free CDN. This data supports the hypothesis that the lectins were preferentially binding to chitin and not *Coccidioides spp*. glycoproteins. Nevertheless, this observation was useful as it spawned the question of whether one could use lectins or other glycan enrichment methods to pull out Coccidioidal glycan antigens from *in vitro* cultures and patient urines. Glycan biomarkers from human and dog urines are evaluated in Chapter V, titled "*Evaluation of Coccidioides-specific glycans from human and dog urines as diagnostic biomarkers for Coccidioidomycosis*."

Finally, Chapter VI, titled "*Discussion*," will discuss the broad impacts of the previous four chapters in improving the diagnostic landscape of Valley Fever in humans and dogs.
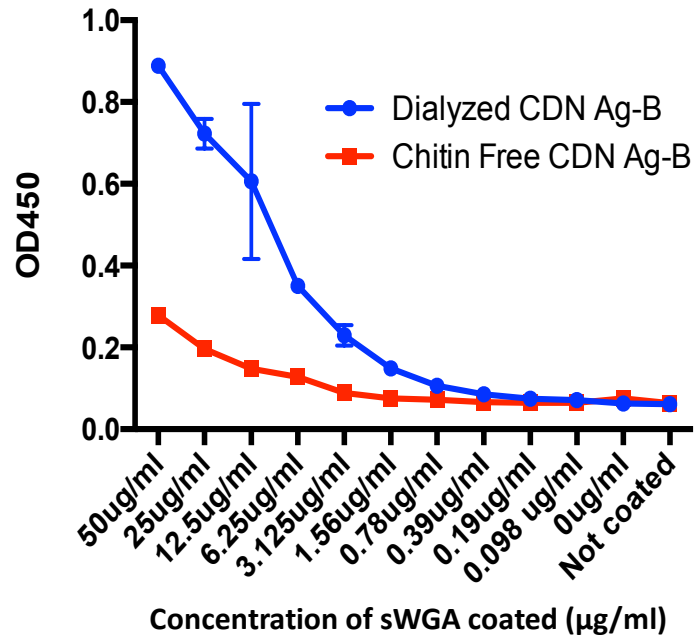
Figure 6. Lectin ELISA shows binding preference to chitin-containing fractions of coccidioidin antigen (CDN Ag). 2μg of Succinylated wheat germ agglutinin (sWGA) was coated to the bottom of a 96 well plate, non-specific sites were blocked with a carbohydrate-free buffer (Vector Labs) and various concentrations of either dialyzed biotinylated CDN Ag (blue) or biotinylated chitin-free CDN Ag (red) were bound prior to washing and binding with a strep-avidin HRP secondary antibody. Error bars represent standard error of triplicate wells.

CHAPTER II

PROTEOGENOMIC RE-ANNOTATION OF

COCCIDIOIDES POSADASII STRAIN SILVEIRA

Natalie M. Mitchell, Andrew L. Sherrard, Surendra Dasari , D. Mitchell Magee, Thomas E. Grys, and Douglas F. Lake

Abstract

The aims of this study were to provide protein-based evidence upon which to re-annotate of the genome of *C. posadasii,* one of two closely related species of *Coccidioides*, a dimorphic fungal pathogen that causes coccidioidomycosis, also called Valley Fever. Proteins present in lysates and filtrates of *in vitro* grown mycelia and parasitic phase spherules from *C. posadasii* strain Silveira were analyzed using a GeLC-MS/MS method. Acquired spectra were processed with a proteogenomics workflow comprising a Silveira proteome database, a 6-frame translation of the Silveira genome and an *ab initio* gene prediction tool prior to validation against published ESTs. This study provides evidence for 837 genes expressed at the protein level, of which 169 proteins (20.2%) were putative proteins and 103 (12.3%) were not annotated in the Silveira genome. Additionally, 275 novel peptides were derived from intragenic regions of the genome and 13 from intergenic regions, resulting in 172 gene refinements.

Additionally, we are the first group to report translationally active retrotransposon elements in a *Coccidioides spp*. Our study reveals that the currently annotated genome of *C. posadasii* str. Silveira needs refinement, which is likely to be the case for many non-model organisms.

Coccidioides immitis and *Coccidioides posadasii* are dimorphic fungal pathogens that cause coccidioidomycosis in a wide range of mammalian hosts, including humans. Mass spectrometry derived proteomic data can help resolve errors in genome annotations and provide empirical evidence for the protein coding genes of microorganisms. Our study reveals that the currently annotated genome of *C. posadasii* str. Silveira is incomplete, and we are able to suggest 172 gene refinements from 288 novel peptides; 58.7% of which had previous EST evidence, but with an additional 41.3% identified by the addition of proteomics data in this study. Furthermore, 74 previously "*C. immitis*-specific" genes were identified in the *C. posadasii* strain, suggesting that proteomically, *C. immitis* and *C. posadasii* are closer than previously annotated. Finally, we are the first group to report a translationally active retrotransposon element in a *Coccidioides spp*.

Introduction

Coccidioides immitis and *Coccidioides posadasii* are dimorphic fungal pathogens that cause coccidioidomycosis in a wide range of mammalian hosts, including humans. *C. posadasii* strain Silveira is one of the most common laboratory strains used in

24

Coccidioides research. Originally isolated in 1951 from a man in the San Joaquin Valley, California, it was initially deposited to the American Type Culture Collection (ATCC) as *C. immitis* Stiles (ATCC 28868), where it continues to be listed as the type strain of the genus, to which all other strains are compared. It was only after 2001 that *C. immitis* and *C. posadasii* were officially designated as two distinct species, based on genomic analyses, and the isolate became *C. posadasii* strain Silveira [69,70].

    C. posadasii strain Silveira fits the archetypical definition of a non-model organism. It requires rigorous category 3 laboratory biosafety conditions, takes 1-3 weeks to grow, and has been poorly-studied, at least in part due to the fact that until 2013, it was a select agent. In 2010, the whole genome of strain Silveira was sequenced (WGS) [71]. To date, transcriptome sequencing (RNA-Seq) has not been performed on this strain, however, Sanger sequencing of cDNA and EST libraries have been performed for Silveira. RNA-Seq has so far only been reported for *C. immitis* RS and *C. posadasii* C735 [72]. As it is well known that ESTs are error-prone and highly redundant [73], and gene transcription levels may not directly correlate with protein production [74-76], a number of questions about Silveira's actual protein-coding genes remain.

    In addition to confirming protein-coding genes, mass spectrometry-based proteomic analyses are increasingly being used in the diagnosis of microbes, identification of drug targets, and in understanding host-pathogen interactions. However, many of the currently annotated proteomes available in protein databases rely heavily on

sequence homology and canonical gene model definitions, rather than evidence at the protein level, especially in non-model organisms like *Coccidioides* species. If proteins are not correctly annotated in the databases, they would be missed in environmental surveys and biomarker discovery studies. In fact, previous proteomic work by our laboratory indicated that several proteins found in proteomic surveys of the parasitic phase of strain Silveira were not annotated in strain Silveira and were missed in the initial analyses [64].

In this study, we performed an in-depth proteomic analysis of strain Silveira to provide empirical evidence of the proteins produced in both the mycelial phase (mycelia) and parasitic phase (spherules). This protein-level evidence was also used to identify novel protein-coding genes, validate previously predicted genes and assist in the refinement of gene models in this organism.

Materials and methods

Culture of mycelia and spherules

Mycelia and spherules of *C. posadasii* strain Silveira were generated as published previously [64,77]. Briefly, mycelial cultures were grown by inoculating liquid culture flasks containing Converse medium with mycelia obtained from three-week-old glucose-yeast extract agar plates. Flasks were continually shaken at 120 rpm at room temperature. For the preparation of spherules, cultures were initiated by seeding flasks of Converse medium with ~ 1-5 X 10^5 arthrospores/ml. Spherules were maintained in continuous

26

culture at 40°C, 20% CO_2, with continuous shaking at 120 rpm in modified Converse medium. At 3 day intervals, the cells from the mycelial and spherule cultures were collected by centrifugation, washed in sterile distilled water, and stored at 4°C in 0.5% formalin. The spent medium was supplemented to 0.5% formalin and was also stored at 4°C until processed, as below.

Processing of proteins from lysates and spent culture medium

Proteins were extracted from the cultures as previously published [78]. Briefly, the spent medium containing proteins expelled during cellular growth (filtrate fraction) was collected for both mycelia and spherules. Proteins were respectively concentrated using a 10,000 MW ultrafiltration membrane. The remaining mycelia and spherules were lysed using mechanical and chemical processes to release intracellular and cell wall proteins (lysate fractions). Spherules were lysed by processing in an ice-cooled beadbeater using 0.5 mm glass beads for 5 minutes. Mycelia were collected and separated from culture filtrate by filtration, collected, and incubated with toluene, at a 3% final concentration in water for 3 days to induce autolysis. Cellular debris was removed by centrifugation (6,000 X g, 10 minutes) and the supernatant was collected, concentrated and lyophilized. The combined filtrate and lysate fractions of the mycelial phase cultures are collectively known as "coccidioidin" and the parasitic phase spherule culture is termed "spherulin." Both coccidioidin and spherulin were lyophilized and stored at -80°C until use.

In-gel digestion of fractions

Three 20 μg technical replicates of coccidioidin and spherulin were heated to 95°C in reducing Laemmli sample buffer (BioRad, Hercules, CA) for 10 minutes prior to separation on 4-20% TGX gels (BioRad). Gels were stained with Bio-Safe Coomassie G-250 Stain (BioRad) as per manufacturer's instructions. Each sample lane of the SDS-PAGE gel was cut into six equal size slices and put into polypropylene tubes prior to destaining twice in water and twice in 50% ACN. Proteins were then reduced in 10mM DTT, and alkylated with 2.5mM ioadacetamide (IAA) prior to 37°C overnight digestion with 200ng trypsin in 25mM ammonium bicarbonate. The peptides were then extracted from the gels using 5% formic acid (FA) in ACN, and dried down in a speed vacuum.

Proteomic Analysis

Protein digests from gel slices were analyzed using LC-MS/MS as previously described [64]. Briefly, each digest was loaded onto a custom made 0.35ul bed OptiPak trap (Optimize Technologies, Oregon City, OR) packed with 5μm, 200Å Magic C18 stationary phase. The loaded trap was washed for 4 minutes with an aqueous loading buffer of 0.2% FA and 0.05% TFA at a flow rate of 10uL/min. Following the wash, peptides were transferred onto a 35cmx100 m PicoFrit column, self-packed with Agilent (Santa Clara, CA) Poroshell 120S 2.7μm EC-C18 stationary phase, using a Dionex UltiMate® 3000 RSLC liquid chromatography system (Thermo-Fisher, Waltham, MA).

Peptides were separated using a 400 nL/min LC gradient comprised of 2%-40% B in 0-70 min. Mobile phase A was 2% ACN in water with 0.2% FA and mobile phase B was ACN/isopropanol/water (80/10/10 by volume) with 0.2% FA. Eluting peptides were analyzed using a QExactive Plus mass spectrometer (Thermo-Fisher). The instrument was configured to operate in data-dependent mode by collecting MS1 data at 70,000 resolving power (measured at m/z 200) with an AGC value of 1E6 over a m/z range of 360-2000, using lock masses from background polysiloxanes at m/z 371.10123 and 446.12002. Precursors were fragmented with normalized collision energy (NCE) of 28, fragments measured at 17,500 resolving power and a fixed first mass of 140. Resulting tandem MS/MS were collected on the top 20 precursor masses present in each MS1 using an AGC value of 1E5, max ion fill time of 50ms, and an isolation window of 1.5 Da.

The provision of information relating to mass spectrometry is in accordance with "Minimum Information About a Proteomics Experiment" (MIAPE) guidelines [79] and the mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [80] with the dataset identifier PXD006489.

Database Construction and Searches

Gene models were confirmed using an annotated protein database generated from the July 14, 2015 release of the *C. posadasii* strain Silveira genome

(www.broadinstitute.org), which is based on WGS data [71,81]. This genome release can also

be found at www.fungidb.org [82]. Proteomic discovery was performed using two

additional databases generated using the genome as source material (Figure 7). The first

was a six-frame translation of the C. *posadasii* strain Silveira genome (WGS assembly

5/31/2007, ABAI02000000; www.ebi.ac.uk), created using the program Sixpack

(www.emboss.sourceforge.net). The second database was generated by the *ab initio* gene

predictor software AUGUSTUS, configured to use *C. immitis* RS as a training set [83]. No

C. posadasii training set was available and this was deemed unnecessary due to the

genomic similarity of the strains and the far superior sequencing coverage of *C. immitis*

RS. Sampling parameters were set at 100 transcripts per locus, exon length at <100 base

pairs. For the MS/MS searches, all three databases were combined and reversed protein

sequences were appended to the database for estimating peptide and protein false

discovery rate (FDR).
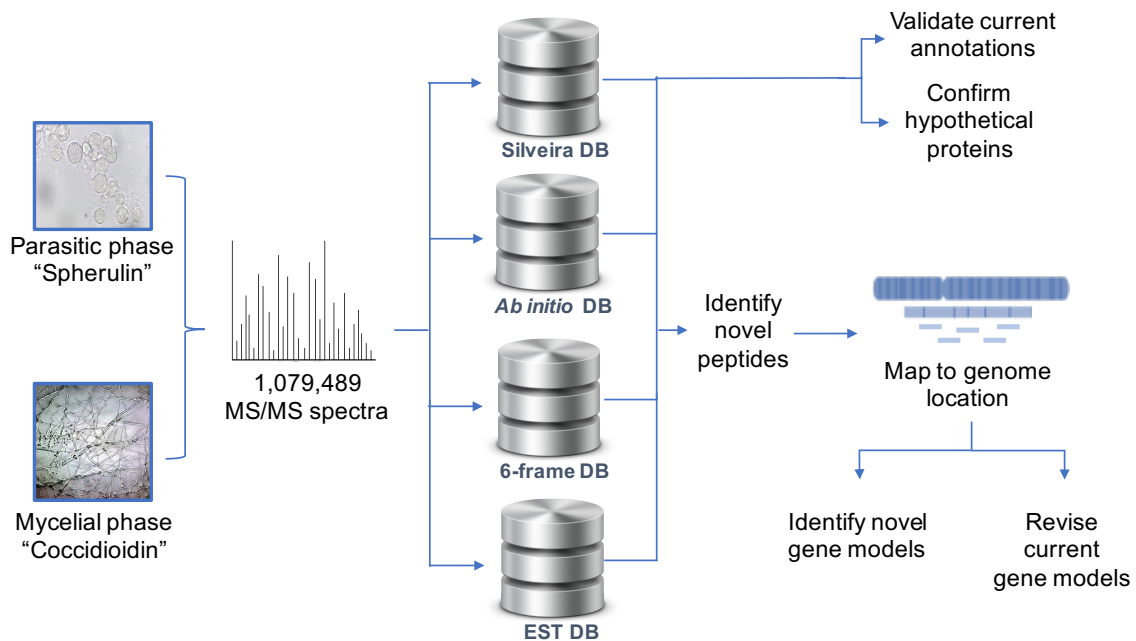
Figure 7. Schematic of the proteogenomics workflow used within this study. Spectral data from mass spectrometry was searched against the currently annotated *C. posadasii* str. Silveira genome, a six-frame translation of this genome and an *ab initio* gene prediction database to validate the current genome annotation, confirm the existence of hypothetical protein, revise current gene models and identify novel peptides. Published ESTs from all available *Coccidioides spp.* were also combined into an EST database, which was used to validate peptides and gene models.

As the genomes of *C. posadasii* and *C. immitis* are very similar, protein sequence databases were also derived from EST data of all *Coccidioides* strains available in the Broad Institute's Coccidioides Genomes project [71,81,84], SwissProt and RefSeq. This EST database was used to validate the peptides and confirm gene models in closely related strains. By considering this EST support from homologous proteins of closely related

Coccidioides species, additional proteins that were not annotated for strain Silveira became identifiable.

Database searching, peptide identification and location

Raw files were converted into mzML format using ProteoWizard's MSConvert tool [85] and the MS/MS spectra were identified using the MS-GF+ software [86] (v2016.06.29) configured to use the aforementioned protein sequence databases. MS-GF+ was also configured to use 10ppm m/z tolerance for both precursors and fragments while performing peptide–spectrum matching and restricted to peptides of length between 7 and 30 amino acids. The software derived semitryptic peptides from the sequence database while looking for the following variable modifications: carbamidomethylation of cysteine (+57.023 Da), oxidation of methionine (+15.994 Da), and formation of n-terminal pyroglutamic acid (−17.023 Da). The resulting identifications were filtered using IDPicker [87] (version 3.1.9729) configured to use MS-GF scoring functions when evaluating the quality of the peptide spectrum matches (PSMs) and return PSMs with <1% FDR. Protein identifications with at least two distinct filtered peptide identifications were reported for further processing.

Data analysis and genome annotation

Mass spectral data were searched against all databases above. Peptides that were present in either the six-frame or the AUGUSTUS databases and absent from the

annotated Silveira protein database were considered novel peptides. Novel peptides were

visually mapped onto the genome using the Integrative Genomics Viewer (Broad

Institute) [88], and peptide homology was determined by tblastn (NCBI) [89]. These novel

peptides were categorized as intragenic if they overlapped an existing gene model or

were within 1000 base pairs of the 5' or 3' ends of a gene model, or intergenic if they did

not. Intragenic peptides were further categorized based on their location relative to the

gene model, in accordance with previously described definitions [73]. Novel peptides were

also matched against the EST databases to provide supportive evidence of their existence.

Transposable element analysis

Putative transposable elements and their associated proteins within the ORFs were

identified using BLAST and long terminal repeats (LTRs) were detected using

LTR_Finder (http://tlife.fudan.edu.cn/ltr_finder/) [90].

Results

The main aims of this study were i) to acquire proteomic evidence of gene

expression in *C. posadasii* strain Silveira in both the mycelial and yeast phases of growth,

ii) to confirm the existence of putative proteins annotated by WGS, and iii) to apply

proteomic evidence to improve the current genome annotation and gene models.

A total of 1,079,489 MS/MS spectra were obtained from 36 GeLC-MS/MS runs and used in the database searches. The 18 runs of the coccidioidin produced 576,300 MS/MS spectra and the 18 runs of spherulin produced 503,189 MS/MS spectra. Spectra were matched to the databases, and filtered with a 1% cumulative FDR at the PSM level and at least two unique peptides per protein. The observed protein-level FDR was 0.64%. The average protein coverage per protein group was 35.1% (range 99.3% to 0.7%) and an average of 12.4 peptides were found per protein (range 2 to 278 peptides; Appendix C).

Validation of the current annotation

We detected a total of 9,024 peptides from 734 proteins present in the annotated Silveira proteome database; 5,315 in coccidioidin and 5,473 in spherulin, with 1,764 found in both. However, when spectra were also searched against the combined EST database created for all other available *C. immitis* and *C. posadasii* strains, an additional 2,491 peptides from 103 proteins were identified, giving a grand total of 837 experimentally identified *Coccidioides* proteins in this study. Of the 103 novel proteins that were not annotated in Silveira, 74 were annotated only in *C. immitis* (8.8%) and 29 (3.5%) were annotated only in non-Silveira *C. posadasii* strains (Figure 8). Thus, 12.3% of the Silveira proteins identified in this study were not annotated in Silveira. Not only does this indicate that the current Silveira genome assembly is fragmented, it provides

evidence that 74 genes that were previously thought to be unique and species-specific to

C. immitis strains are also expressed in *C. posadasii*.

■ Annotated to *C. posadasii* str. Silveira
☐ Not annotated to *C. posadasii* str. Silveira
■ Annotated to *C. immitis* strains only (RS / RMSCC 2394 / RMSCC 3703 / H538.4)
■ Annotated to *C. posadasii* strains other than Silveira only (C735 / RMSCC 3488)
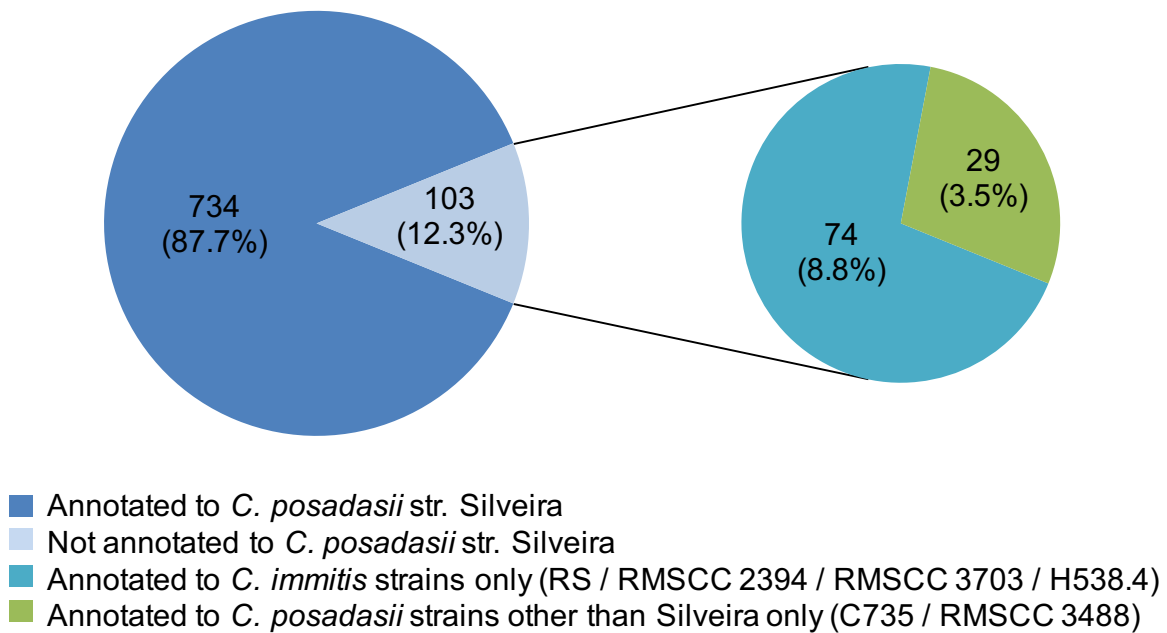
Figure 8. Number of protein groups identified and the *Coccidioides* species to which they were annotated suggest *C. posadasii* and *C. immitis* are closely related at the protein level. There were 734 proteins that were correctly annotated to *C. posadasii* str. Silveira and 103 novel proteins that were not correctly annotated to the strain. Of the 103 novel proteins for *C. posadasii* str. Silveira, 74 (8.8%) were only annotated to *C. immitis* strains, and 29 (3.5%) were annotated to other *C. posadasii* strains but not strain Silveira.

Hypothetical protein identification

Hypothetical and putative uncharacterized proteins account for 5,073/10,212 (49.7%) of the currently annotated proteins in strain Silveira (by WGS). This indicates that prior to this study, approximately half of all genes annotated for strain Silveira had no empirical evidence to corroborate that they truly are genes or are functional. In this study, 146 (19.9%) of the proteins identified and annotated in Silveira were previously classified as hypothetical or putative proteins. This study thus gave experimental evidence for the expression of 2.9% of the total hypothetical/putative proteins annotated in Silveira.

Novel peptide identification

We identified a total of 288 novel peptides that did not conform to a known gene model in strain Silveira (Appendix B). Of these novel peptides, 275 (95.5%) were mapped to intragenic regions and 13 (4.5%) to intergenic regions (Table 1). A total of 199 novel peptides (58.7%) had previous EST evidence, however, an additional 119 novel peptides (41.3%) were identified by the addition of proteomics data in this study. Through the identification and mapping of these novel peptides using the 2 peptides per protein rule, we are able to suggest 172 gene refinements to the Silveira genome. Some examples of these gene refinements are shown in Figure 9; an example of a peptide found within a sequence gap (Figure 9a), peptides in additional exons (Figure 9b), a peptide

36

located within a 5' UTR (Figure 9c), and a peptide crossing an intron-exon boundary

(Figure 9d). There were 17 peptides found to cross intron-exon boundaries and 10

peptides found out of frame of the existing annotation. Of the 13 intergenic peptides, all

were identified using EST evidence from other *Coccidioides spp*., and all were found to

reside within the gaps between contigs. This illustrates that strain Silveira requires better

genomic sequencing coverage than is currently available.

Category of Modification	Count	# with EST Evidence	# without EST Evidence
Intragenic Peptides	**275**	**156**	**119**
Peptides mapped to intronic regions	0	-	0
Peptides out of frame of existing annotation	11	9	2
Peptides crossing intron-exon boundaries	17	12	5
Peptides mapped to 5' UTRs	2	0	2
Peptides mapped to 3' UTRs	0	0	0
Peptides part of putative transposons	3	0	3
Peptides outside of annotated Contig (in gaps)	4	4	0
Novel exon	107	57	50
Single nucleotide polymorphisms	73	73	0
Other peptides too short for confident description	58	1	57
Intergenic Peptides	**13**	**13**	**0**
Total Novel Peptides	**288**	**169**	**119**

Table 1. Summary of novel peptides identified by LC-MS/MS in mycelial and parasitic phase fungi. There was a total of 288 novel proteins identified in this study. There were 275 novel intragenic peptides, with 156 of these (56.7%) having EST support from other *Coccidioides spp*. An additional 13 novel peptides were located within intergenic regions, all of which (100%) had EST support from other *Coccidioides spp*. The 119 novel peptides that did not have EST evidence indicate that the addition of proteomics data in this analysis allowed for additional gene refinements beyond that which was provided by EST evidence alone.
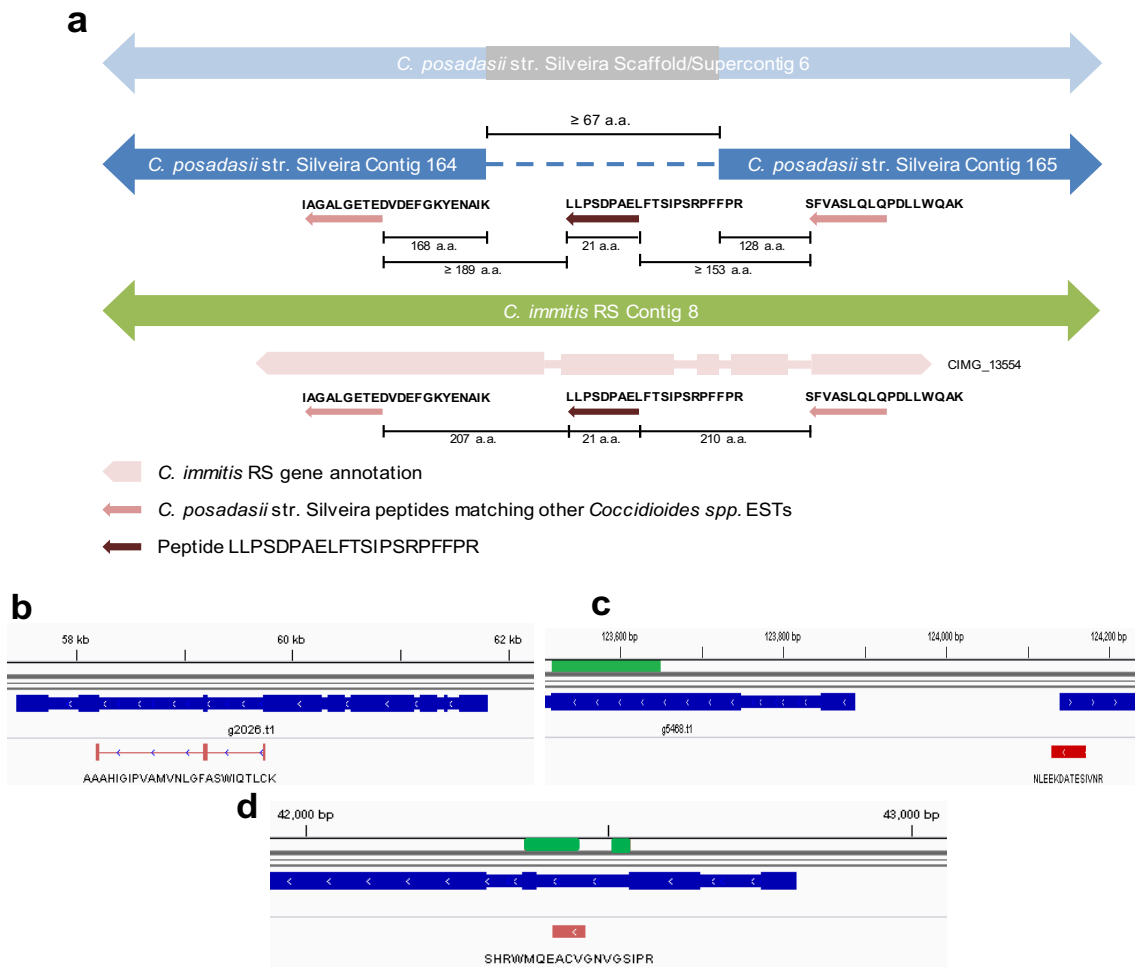
Figure 9. Examples of some observed intragenic events suggest re-annotation. a) Peptide LLPS… was found to be an additional exon located in the gap between contigs 164 and 165. This peptide was not found in the Silveira 6-frame database and required mapping to the *C. immitis* RS genome in Integrative Genomics Viewer (IGV) to be located; **b)** IGV snapshot of peptide AAAH… crossing three additional exons of CPSG_02828 in contig 103; **c)** IGV snapshot of peptide NLEE… in the 5' UTR region of CPAG_0201, contig 290; **d)** IGV snapshot of peptide SHRW…crossing an intron/exon boundary in CPSG_03385, contig 133. Blue rectangles and boxes represent AUGUSTUS predictions of introns and exons, green rectangles are annotated exons and red rectangles are the peptides identified by mass spectrometry, with their relative location on the contig.

Putative transposon

A cluster of 3 peptides (QSQWDKDQNEK, GASQQDNEGGYPR and TPVKPLPK) were found in 6 different locations through the length of the genome (Figure 10a) and did not share sequence homology to proteins in any other *Coccidioides* species or have EST support. Due to the multiple genomic locations of this peptide cluster, we hypothesized that this is part of a transposable element (TE). Since multiple spectra (n=6) of this peptide cluster were found, some elements of this putative TE must be translationally active in at least one of the six locations. Potentially, it has inserted itself adjacent to a promoter or replaced/displaced a normally expressed gene. When searched in LTR_Finder, intact LTRs were found in Contig 370 and Contig 368, aiding the classification of this TE as a retroviral element. Since LTRs are retroviral promoters, this supports the hypothesis that the retroviral elements are translationally active in at least one of these locations. When additionally searched in NCBI's Conserved Domain Search, Contig 368 was found to also contain a number of retrotransposon elements of *gag* and *pol* proteins. It was found that in addition to the LTR, retropepsin, reverse transcriptase, RNaseH, integrase and a Chromovirus domain were found (Figure 10b). Based on the sequence of the protein domains, it is likely that this retrotransposon is from the Ty3/Gypsy superfamily, and because it contains a chromodomain in the *pol* ORF, it is likely a Chromovirus. No stop codons were found around the *gag* proteins; however,

40

there were stop codons in the *pol* region. It is therefore likely that only the *gag* proteins are being translated and the retrotransposon is not transpositionally active.

a

Contiguous Sequence	Frame	ORF
Contig 79	3	3099
Contig 114	2	188
Contig 368	-3	365
Contig 370	1	2630
Contig 370	2	511
Contig 420	3	190

b
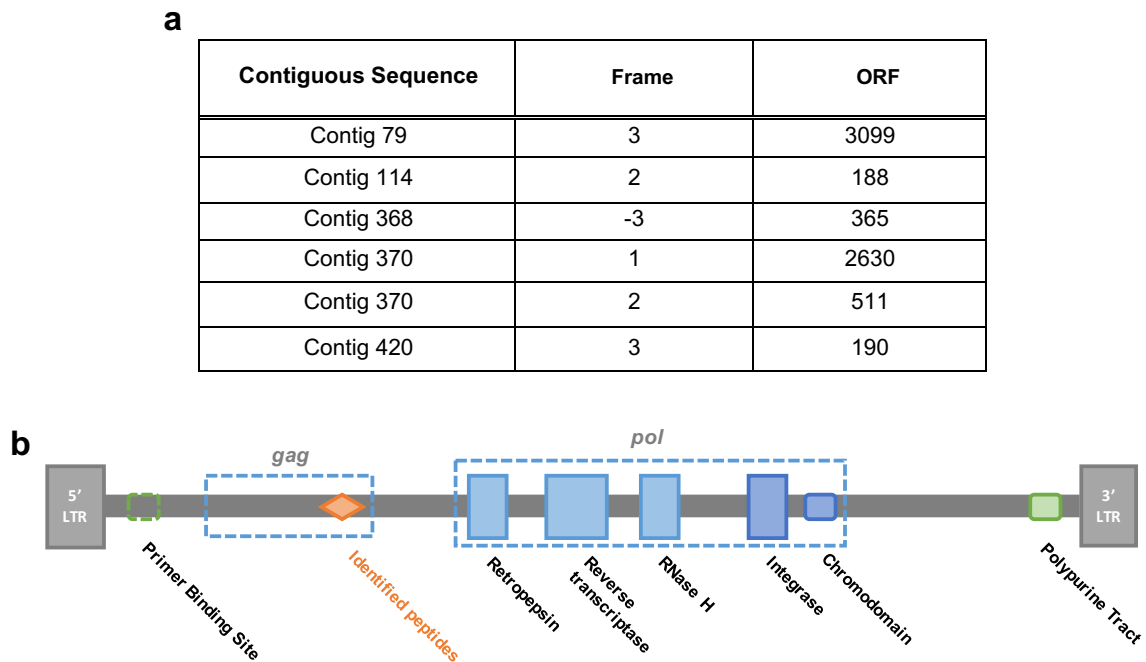
Figure 10: Putative transposon with actively translated retroviral elements located in six ORFs. a) Genomic locations of recurring retrotransposon peptides QSQWDKDQNEK, GASQQDNEGGYPR and TPVKPLPK. This cluster of peptides was found to be located in six different ORFs, on six different contigs, in four different frames. **b)** Diagram of retroviral transposon *gag* and *pol* protein elements identified near the novel peptides QSQWDKDQNEK, GASQQDNEGGYPR and TPVKPLPK on Contig 368, ORF 365. These peptides were found upstream of a retropepsin, reverse transcriptase, RNaseH, integrase and Chromovirus domain, with flanking LTR elements.

41

Discussion

C. *posadasii* strain Silveira is one of the most studied laboratory strains of *Coccidioides*, yet the genome sequence coverage is incomplete. It is one of only two *C. posadasii* strains available for purchase from the ATCC and is the listed type strain of the genus. Therefore, it is important that this strain is properly annotated. In this study, we employed mass spectrometry to identify the proteins produced by strain Silveira and then used these proteins to revise its current genome annotation. Through the use of proteogenomics, we have identified 172 genes in need of refinement and provided evidence for 837 proteins produced by Silveira, of which 103 had not been previously annotated in Silveira. We maximized the identification of expressed proteins by analyzing both the mycelial and parasitic phases of the fungus, with 5,315 peptides found in the mycelial phase and 5,473 peptides identified in the parasitic phase. Additionally, we confirmed the existence of 146 hypothetical and putative proteins for which there was no previous evidence. As hypothetical and putative genes are predicted based on sequence and gene-like context, it is our hope that this report will stimulate studies to characterize the biological role played by some of these proteins discovered here. Additional proteins are likely to be empirically confirmed and characterized as more sensitive mass spectrometers are developed and more investigators integrate high resolution mass spectrometry into their research.

The number of proteins identified in this study (n=837) represents 8.20% of the 10,212 genes currently annotated for the strain. Similar shotgun proteomics experiments in the well-studied fungal strain *Aspergillus fumigatus* Af293 (29Mb genome size), which has 9,648 annotated genes, only identified 414 proteins in the mycelial form. Hence, the proteomic depth sampled in this study compares favorably to proteogenomic studies in other fungi [91]. RNA sequencing data from mycelial and parasitic phases of *C. posadasii* strain C735 was able to quantify several thousand transcribed genes [72]. Many of these mRNAs, even if transcribed with a five-prime cap and polyadenylated tail, could be unstable, rapidly degraded or repressed by any one of a number of RNA interference pathways (RNAi) that have been identified in fungi [92]. It is known that fungi produce a diverse range of small noncoding molecules like microRNAs (miRNAs) and short interfering RNAs (siRNAs) which can silence or degrade mRNA transcripts, preventing their translation into a fully functional protein [93,94]. Future experiments that perform tandem RNA sequencing or RT-PCR and mass spectrometry on the same biological sample could help elucidate how many transcribed mRNAs become translated into a protein.

Our data also showed that the majority of novel genes identified in *C. posadasii* str. Silveira (74/103) were annotated only in *C. immitis* (Figure 8). This indicates that these 74 genes are also missing from the annotations for *C. posadasii* strains C735 and RMSCC3488; indicating that these genomes are also in need of revision. Interestingly,

when the current Silveira genome was insufficient to locate novel peptides in this study,

we were able to map them to their homologous genes by mapping to the *C. immitis* RS

genome in IGV. In addition to highlighting the need for reannotating the genome, this

data suggests that proteomically, *C. immitis* and *C. posadasii* are more closely related

than previously annotated; as a number of *C. immitis* "species-specific" genes were also

found in *C. posadasii* strain Silveira. Although molecular methods have been developed

that speciate *C. posadasii* and *C. immitis*, it is unclear how many functionally different

proteins are expressed by one but not the other species. It is estimated that the

nonrepetitive genomic portions of *C. immitis* and *C. posadasii* strains are 93.5%

homologous, with a median sequence identity of 98.3% [81]. Using FungiDB [82], a

functional genomics fungal database to search for the current list of *C. posadasii* unique

genes that are not found in an *C. immitis* strains, a genome phylogeny comparison

indicated that there are only 18 *C. posadasii*-unique genes in *C. posadasii* C735, whereas

there are 701 in *C. posadasii* RMSCC 3488 and 821 in *C. posadasii* str. Silveira. In this

study, we only had protein evidence of one of the 821 *C. posadasii*-unique genes

(CPSG_01866, a hypothetical protein), lending to the possibility that at least some of

these genes may be artifacts of sequencing errors, read gaps, or errors in contig assembly.

Both *C. posadasii* trains RMSCC 3488 (287 contigs, N50=235kb, assembly gap length=

120kb) and Silveira (464 contigs, N50=140kb, assembly gap length=160kb) have much

greater numbers of contigs with ten-fold lower N50 contig quality scores and much

higher assembly gap lengths than strain C735 (58 contigs, N50=2,400kb, gap length=0), and *C. immitis* RS (11 contigs, N50=4,300kb, gap length=400bp), lending some credence to this possibility. Regardless, since infection with either *C. posadasii* or *C. immitis* appear to result in the same clinical disease, further work in this area is warranted and studies that evaluate the proteomes of multiple Coccidioides spp. strains simultaneously prior to proteogenomics evaluation would be very enlightening.

Transposons are important elements of genome plasticity and have been previously used to identify species-specific variations and evolutionary divergence [71]. Infectious agents that cause systemic disease, like *Coccidiodies* species, are linked with elevated LTR retrotransposon content, whereas the closely related non-infectious agent *Uncinocarpus reesei* has low LTR retrotransposon content [95]. This has been attributed to the fact that actively mobile elements indicate a strong need for adaptation to new environments, like those encountered within different hosts and ecological niches. It is possible that the high rate of transposon integration into the geographically diverse *Coccidioides* species has pushed greater genomic diversity of the strains, without creating significant proteomic diversity.

Within this paper, we identified a putative transposon that was found at six locations throughout the genome. As we identified three peptides from within this transposon by mass spectrometry, at least some of the elements must be translationally active in at least one of these genomic locations. This is the first report in the literature

45

for protein evidence of an actively translated retrotransposon in any *Coccidioides spp*. It has been previously stated that proteomic analyses alone are not sufficient to identify novel transposons and the addition of transcriptional data is required to do so adequately [96]. Although we did use ESTs to provide additional supportive evidence for other novel peptides in this study, there were no ESTs to support the three peptides of this transposon. However, we were still able to verify the existence of this novel retrotransposon as the peptides matched to the six-frame translation of the genome and to a gene prediction by AUGUSTUS. When these peptides were mapped to the genome in IGV, the visualization that this peptide cluster was on six separate contigs led us to identifying multiple retrotransposon element motifs in BLAST and LTR Finder. We were able to identity this putative novel transposon as belonging to the Ty3/Gypsy clade; likely a Chromovirus due to the chromodomain. This is consistent with previous publications that determined that the majority of transposons identified in *Coccidioides* genomes were of this clade [71].

Reannotion of the Silveira genome is important for researchers pursuing diagnostic markers or identifying protein expression. The additional genes identified in this study may be included in future database updates. Without such a reannotation of the Silveira genome, researchers could be blind to observing certain proteins in a sample, purely because the proteins are not available in the given databases. A combination of both improved gene-prediction algorithms/software and re-sequencing to a greater

46

coverage with newer sequencing technology that offers longer reads, is likely required to improve the accuracy of the genome of *C*. *posadasii* strain Silveira. However, a useful first step in the interim would include a proteomic evidence track and a transcriptomic evidence track in genomic resource websites like FungiDB. This would allow researchers to easily identify annotation errors, so that erroneous data are not used in subsequent data analysis, further propagating data inaccuracies.

CHAPTER III

LCM-ASSISTED PROTEIN BIOMARKER DISCOVERY FROM *COCCIDIOIDES*-INFFECTED LUNG TISSUE

Abstract

Laser capture microdissection (LCM) coupled to label-free quantitative mass spectrometry is a strategy to identify the most probable biomarkers from archival tissues. In this study, LCM was used to take a "snapshot" of proteins produced *in vivo* during *Coccidiodies spp*. infection in human, mouse and dog hosts. From the 100 most abundant proteins produced *in vivo*, 28 *Coccidioides spp*. biomarker candidates were identified which do not share >40% sequence orthology with any human protein. Three of these biomarker candidates are also potential *Coccidoides*-specific biomarkers, as they also do not share sequence homology to any other pathogenic fungus or microbe. Gene ontology analysis of the 28 biomarker candidate proteins revealed enriched hydrolase activity and increased purine and carbohydrate metabolism functions. Additionally provided is proteomic evidence that all 28 biomarker candidates are produced by the fungus when grown *in vitro*, in a media- and growth-phase dependent manner.

Introduction

A specific and sensitive *Coccidioides spp*. antigen released from spherules in the acute phase of Valley Fever infection (VF) is the holy grail of VF diagnostics. Detection of antigen in a body fluid would allow for a confident and early clinical diagnosis of VF; something that is rarely attainable with current diagnostic methodologies. A *Coccidioides*-specific protein antigen found in either blood or urine is yet to be identified for VF. With over 3,500 plasma proteins in the Human Proteome Organization (HUPO) human plasma proteome database[97,98] and an estimated >500 proteins circulating at any one time in human plasma,[99] identification of low abundance biomarker proteins among high abundance plasma proteins like albumin and immunoglobulins is a difficult challenge in mass spectrometry-based biomarker discovery.[99,100] To be able to identify low abundance proteins, enrichment using selective antibodies or other affinity reagents are frequently employed options. However, in order to identify the correct affinity reagent or to produce a selective antibody, one must know the target protein biomarker.

Identification of infectious disease proteins from pure *in vitro* cultures of a microorganism is not difficult. However, the most abundant proteins produced during *in vitro* culture may not be fully representative of the *in vivo* proteome of organisms. Protein production is known to be influenced by a number of host pressures not available *in vitro* (immune cells, adhesion, hypoxia, nutrient deprivation, osmotic stress, etc.). Significant differences in phenotype, virulence and survival have thus been noted between fungi

49

grown *in vivo* versus *in vitro*.[101,102] Additionally, even different *in vitro* culture media formulations can cause significant phenotypic changes in *Coccidioides spp*.[103]

One biomarker discovery platform which is able to evaluate protein abundance *in vivo* is laser capture microdissection (LCM) followed by mass spectrometry. LCM facilitates the sampling of specific cellular subtypes from *ex vivo* tissues in a reasonably high throughput manner. In this chapter, an LCM-assisted label-free proteomics pipeline for the identification of VF protein biomarkers from 3 different mammalian host lung tissues is described. Protein abundances from *in vivo s*pherules are then compared to the protein abundances produced from *in vitro*-grown mycelia and spherules in different media.

Materials and methods

Coccidoides spp.-infected tissue samples

Formalin-fixed paraffin embedded (FFPE) lung tissue blocks used in this study were selected from culture-confirmed cases of *Coccidioides spp*. infection from three mammalian species. Triplicate technical replicates of lung tissues from 3 naturally-infected human clinical cases (2 HIV positive individuals and 1 immunosuppressed with Humira), a laboratory-inoculated BALB/c mouse given *C*. *posadasi* strain Silveira (ATCC 28868) via aerosol infection, and a naturally infected dog from a veterinary practice, were performed. Human tissue blocks were acquired from Mayo Clinic Arizona

histology biobank in accordance with IRB# 2-000965 (human tissue). Mouse and dog lung tissue blocks were kindly provided by Dr. Lisa Shubitz of the Valley Fever Center of Excellence. Ten micrometer thick sections were cut using a microtome and transferred to 1mm polyethylene naphthalate (PEN) membrane slides (Carl Zeiss Microscopy; Jena, Germany), deparaffinized and stained with hematoxylin and eosin stain (H&E) using standard procedures.

LCM of FFPE tissues

Laser capture microdissection (LCM) was performed using a Zeiss PALM MicroBeam scope with RoboPalm software (Carl Zeiss Microscopy). Approximately 500,000 μm^2 area of tissue was collected for each sample by laser capture. Tissue features were selected, collected and catapulted into the cap of 0.5 ml Eppendorf tubes containing 35 μl of 0.1 M Tris-HCl (pH = 8.0), 0.002% Zwittergent Z3-16 (MilliporeSigma) via laser pressure catapulting (Figure 11). Immediately after capture, the tube was centrifuged at 14000×g for 2 min. to collect the lysis solution and tissue into the bottom of the tube and was frozen at -80C until processing.
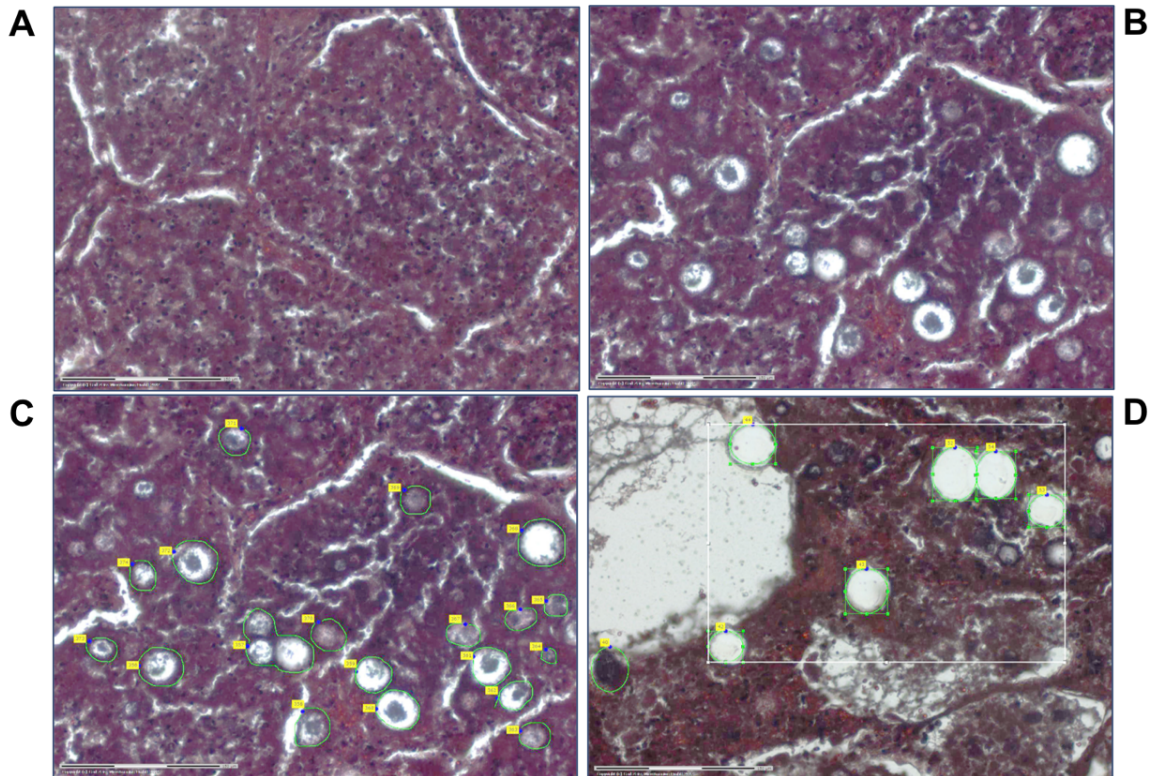
Figure 11: Laser capture microdissection (LCM) of *Coccidioides spp.*-infected human lung tissue is able to precisely extract spherules. Clear (white) circles rimmed with green and with yellow boxes indicate areas where spherules were laser-captured. Tissue sections without fungal spherules (A) can be ignored, whereas sections with spherule elements (B) can be selected (C) and exclusively removed (D).

In-solution protein digestion

Formalin-protein crosslinks were broken from the tissue fragments by heating the sample at 99°C for 1 h in a thermalcycler. Proteins were reduced by the addition of 1.8μl of 200mM tris(2-carboxyethyl)phosphine (TCEP, MilliporeSigma) in 0.1mM Tris, pH 8.2, (final conc. TCEP = 10mM) and incubated at 56°C for 30 min. Afterwards, proteins

were alkylated by the addition of 1.9 μl of 200mM iodoacetamide (IAA, MilliporeSigma) in 0.1 M Tris-HCl (final conc. = 10 mM), and incubated in the dark at RT for 30 min. Finally, proteins were digested by adding 2 μl of 0.01 μg/ul trypsin (Thermo) and incubating at 37°C for 16 h. The digestion was terminated by adding 4 μl of 2% formic acid (FA) prior to loading on the mass spectrometer. Loading was normalized to the area of tissue, optimally 500,000 square microns per sample provides ~500 ng protein.

Fungal culture preparation and protein lysis

The mycelial form of *Coccidioides posadasii* strain Silveira (strain kindly provided by Dr. Bridget Barker of the University of Arizona) was grown in a shake flask rotating at 180RPM in 30°C for 7 days in one of the following six liquid media: 1) Levine modification of Converse media[104] with 0.5% N-Tamol (Dow Chemical Company, Midland MI), 2) RPMI-1640 medium (Corning) with 10% fetal bovine serum (FBS) or 3) RPMI-1640 medium with 0.1% Survanta (Beractant bovine lung surfactants; Abbot Laboratories, Chicago IL).

Spherule form *C. posadasii* strain Silveira cultures were grown in vented shake flasks rotating at 180RPM in 20% CO_2 at 40°C for 7 days in the same three media as above. Cultures were centrifuged at 4,500RPM for 30min. in a Thermo Forma Multi RF tabletop centrifuge (Thermo). Supernatants were then sterilized by filtration through 0.45μm vacuum filtration units (MilliporeSigma). Culture pellets were resuspended 5:1

53

in fungal lysis buffer (50 mM Tris, pH 7.6, 100 mM NaCl, 50 mM EDTA, 5% SDS) prior to transfer into 1.5ml silicone washer-sealed internally-threaded cryovials containing ~350μl of 0.5mm acid washed sterile glass beads. Cryovials were subjected to two rounds of 15 min. bead beating at maximum speed on a vortex with 1.5ml tube attachment unit followed by 3 cycles of flash freezing on dry ice. Sterility was checked by plating 10% of the total volumes onto 2x glucose yeast extract agar (2×GYE) for 3 weeks at RT.

Fungal culture protein extraction and in-gel protein digestion

Fungal culture proteins from filtered supernatants and pellet lysates were extracted using a modification of a previously published protocol.[105] In brief, filtrates and lysates were centrifuged for 30min. at 8,000RPM at 4°C. The supernatants were put into fresh tubes and proteins were precipitated by the addition of 4 volumes of ice cold 10%w/v trichloroacetic acid (TCA, MilliporeSigma) in acetone with 0.007% w/v dithiothreitol (DTT, G Biosciences, St. Louis, MO). Samples were centrifuged at 3,000 g for 10min. and the resultant protein pellets were washed three times in ice cold acetone with 0.007% DTT. The final pellet was then resuspended in rehydration buffer [7M urea, 2M Thiourea, 4% w/v 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate (CHAPS, all MilliporeSigma) and 20mM DTT].

The samples in rehydration buffer were then filter exchanged with phosphate

buffered saline (PBS), using Amicon® Ultra 0.5ml 3kDa filtration units

(MilliporeSigma) prior to protein estimation using a micro bicinchoninic acid (BCA)

protein assay kit (Thermo). ~10μg of protein from filtrate samples and ~30μg of

corresponding lysate protein were mixed with reducing Laemmli sample buffer (BioRad,

Hercules, CA), heated at 95°C for 5min. prior to loading on mini-Protean™ TGX precast

gels(BioRad). Following electrophoresis, gels were stained with Bio-Safe Coomassie G-

250 Stain (BioRad) as per manufacturer's instructions. Each sample lane of the SDS-

PAGE gel was cut into six equal size fractions. Band fractions were then further reduced

into cubes of 1–2mm^3, put into low protein binding tubes (Eppendorf, Hamburg,

Germany) prior to destaining. Proteins were reduced in 10mM DTT for 30 min. at 60°C,

and alkylated with 55mM ioadacetamide (IAA, MilliporeSigma) for 30min. at room

temperature in the dark, prior to 37°C overnight digestion with Pierce™ MS Grade

trypsin protease (Thermo) diluted to 20ng/ml in 100mM ammonium bicarbonate

(MilliporeSigma). The peptides were then extracted from the gel pieces using 5% FA,

ammonium bicarbonate and acetonitrile washes prior to being dried in a speed vacuum

and stored at -80°C until LC-MS analysis.

Proteomic analysis of fungal lysates and filtrates

Protein digests were reconstituted in 0.1% FA and analyzed using LC-MS/MS by loading onto a Dionex UltiMate® 3000 RSLC liquid chromatography (LC) system (Thermo, San Jose, CA) using a PepMap RSLC C18 2um, 75um x 50cm EASY-Spray™ column (Thermo). Peptides were separated using a 0.3 uL/min LC gradient comprised of 2%-90% mobile phase B in 0-120 min. Mobile phase A and B were 0.1% FA in water and acetonitrile, respectively. Eluting peptides were directly injected into an Orbitrap Elite Velos mass spectrometer (Thermo) and ionized using collision-induced dissociation (CID) in positive ion mode. A "top 15″ data-dependent MS/MS analysis was performed (acquisition of a full scan spectrum followed by collision-induced dissociation mass spectra of the 15 most abundant ions in the survey scan).

Protein Identification and label-free protein quantification

Database searching was performed using Sequest (Thermo) in Proteome Discoverer v1.4.1.14 (Thermo) against a combined FASTA database of all the most recent Uniprot *Coccidioides spp*. proteomes (*Coccidioides immitis* RS, proteome ID: UP000001261, June 12, 2018 release date; *Coccidioides immitis* RMSCC 3703, proteome ID:UP000054559, February 26, 2018 release date; *Coccidioides immitis* H538.4, proteome ID: UP000054563, February 26, 2018 release date; *Coccidioides immitis* RMSCC 2394, UP000054565, February 26, 2018 release date; *Coccidioides posadasii*

strain RMSCC 757 / Silveira, proteome ID: UP000002497, February 26, 2018 release date, *Coccidioides posadasii* C735, proteome ID: UP000009084, November 9, 2018 release date; and *Coccidioides posadasii* RMSCC 3488, UP000054567, October 26, 2018 release date). Searches were performed using a fragment tolerance of 0.60 Da (Monoisotopic), parent tolerance of 10 ppm (Monoisotopic), with carbamidomethyl of cysteine as fixed and oxidation of methionine as variable modifications with maximum missed cleavages allowed of 2. Protein identifications were accepted if they achieved a minimum of 2 peptides per protein and a false discovery rate (FDR) of <1%. Label-free protein quantification was performed using normalization of spectral abundance factors (NSAF) in Scaffold (v4.8.7, Proteome Software Inc.).

Biomarker identification and bioinformatics

The top 100 most abundant *Coccidioides spp*. proteins found in LCM lung tissues from the mean of triplicate technical replicates of human, mouse and dog tissues were identified. These "top 100" most abundant coccidioidal proteins were subjected to a pBLAST search (www.blast.ncbi.nlm.nih.gov) against the Homo sapien proteome (taxid:9606), to exclude human orthologues.

Gene ontology identifiers of each remaining protein were pulled from Uniprot (www.uniprot.org). A gene ontology enrichment scatterplot was produced using the Revigo[106] plugin in FungiDB (www.fungidb.org).[107] The wordcloud of enriched KEGG

metabololic pathways was produced using the GOSummaries[108] plugin, also from FungiDB. Protein O-glycosylation predictions were made using NetOGlyc server 4.0.[109] N-glycosylation and signal peptide predictions were determined using NetNGlyc 1.0 server.[110]

Statistical analyses

GraphPad Prism v.8.0 was used for production and statistical analyses, including those used in creating volcano plots and box and whisker plots. Volcano plots were created using multiple t tests of Normalized Spectral Abundance Factor (NSAF), corrected for multiple comparisons using Holm-Sidak method. NSAF is a unitless, arbitrary value used to rank abundance of proteins across samples and experiments.[111] Analysis of *in vivo* versus *in vitro* grown spherule abundances was calculated using a one-way ANOVA followed by Tukey test and adjusted for multiple comparisons using Sidak's method.

Results

Proteomic analysis and biomarker selection

There were a total of 807 *Coccidioides-spp*. proteins identified from the combined human, dog and mouse lung tissue samples used in this study by mass spectrometry. Using a 1% FDR cut-off, there were 777 proteins identified in mouse, 326 from human,

and 123 from dog lung tissues (Figure 12). Ninety-six proteins were common to all three host species and 323 proteins were common to at least 2 host species. The 807 *Coccidioides-spp*. proteins were then mined for possible biomarkers. The mean NSAF of triplicate technical replicates from human, mouse and dog lungs were calculated and the top 100 most abundant proteins were identified.

Seventy-two *Coccidioides spp*. proteins with human orthologues (pBLAST $\geq$40% identity) were removed from the dataset, leaving 28 possible *Coccidioides spp*. biomarkers. All 28 biomarker candidates were found in at least 2 host species lung tissues. Only 2 proteins found in *Coccidioides*-infected mouse tissues were missing from *Coccidioides*-infected human tissue(Ag2/PRA and Uncharacterized protein CISG_09979) and only 1 protein was missing from *Coccidioides*-infected mouse lung tissue which was found in *Coccidioides*-infected human tissue (Flavodoxin domain containing protein).

A pBLAST search of the 28 possible biomarker proteins against all fungi (taxid: 4751) revealed between 53% and 87% shared identity with other pneumonia-causing fungal species for 25/28 of the biomarker candidates. Only 3 proteins did not share significant identity with other pneumonia-causing fungal species; of which all 3 of these proteins were uncharacterized proteins (CIMG_00509, CIMG_05576, CIMG_09001), and shared sequence identity in *C. immitis* and *C. posadasii* strains.
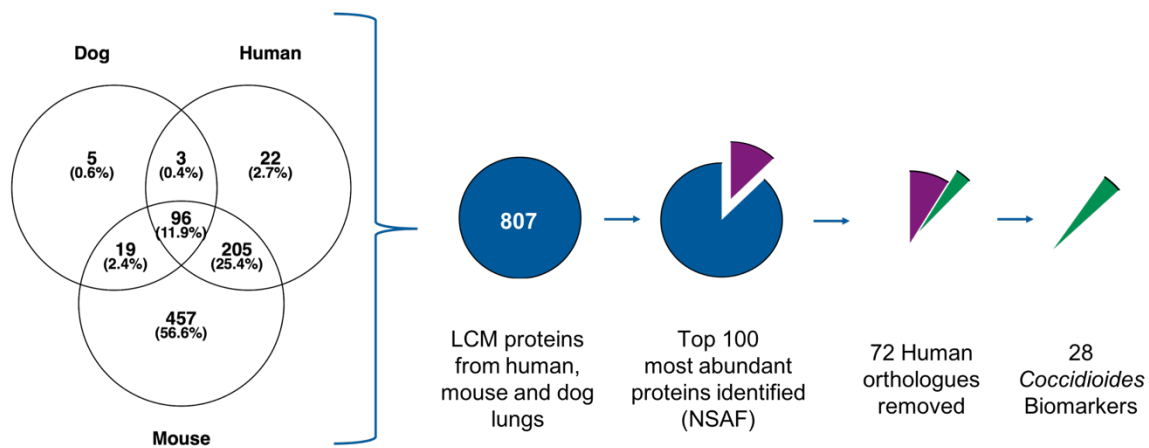
Figure 12. Laser capture microdissection (LCM) assisted biomarker discovery flowchart. There were a total of 807 *Coccidioides spp*. proteins identified in infected lung tissues by mass spectrometry. Proteins with the 100 highest normalized spectral abundance factors (NSAFs) were further evaluated for orthology to human proteins (pBLAST ≥40% identity). Seventy-two human orthologues were removed, leaving 28 *Coccidioides spp*. biomarker candidates.

Gene ontology enrichment analysis of the 28 biomarker candidates indicate significant increases in certain molecular functions and metabolic pathways. As seen in Figure 13A, catalytic, hydrolase and oxidoreductase activities are most increased, with moderate increase in carbohydrate derivative binding. Enrichment of KEGG metabolic pathways in Figure 13B indicate increased purine/pyrimidine metabolism, as well as glycolysis/gluconeogenesis.
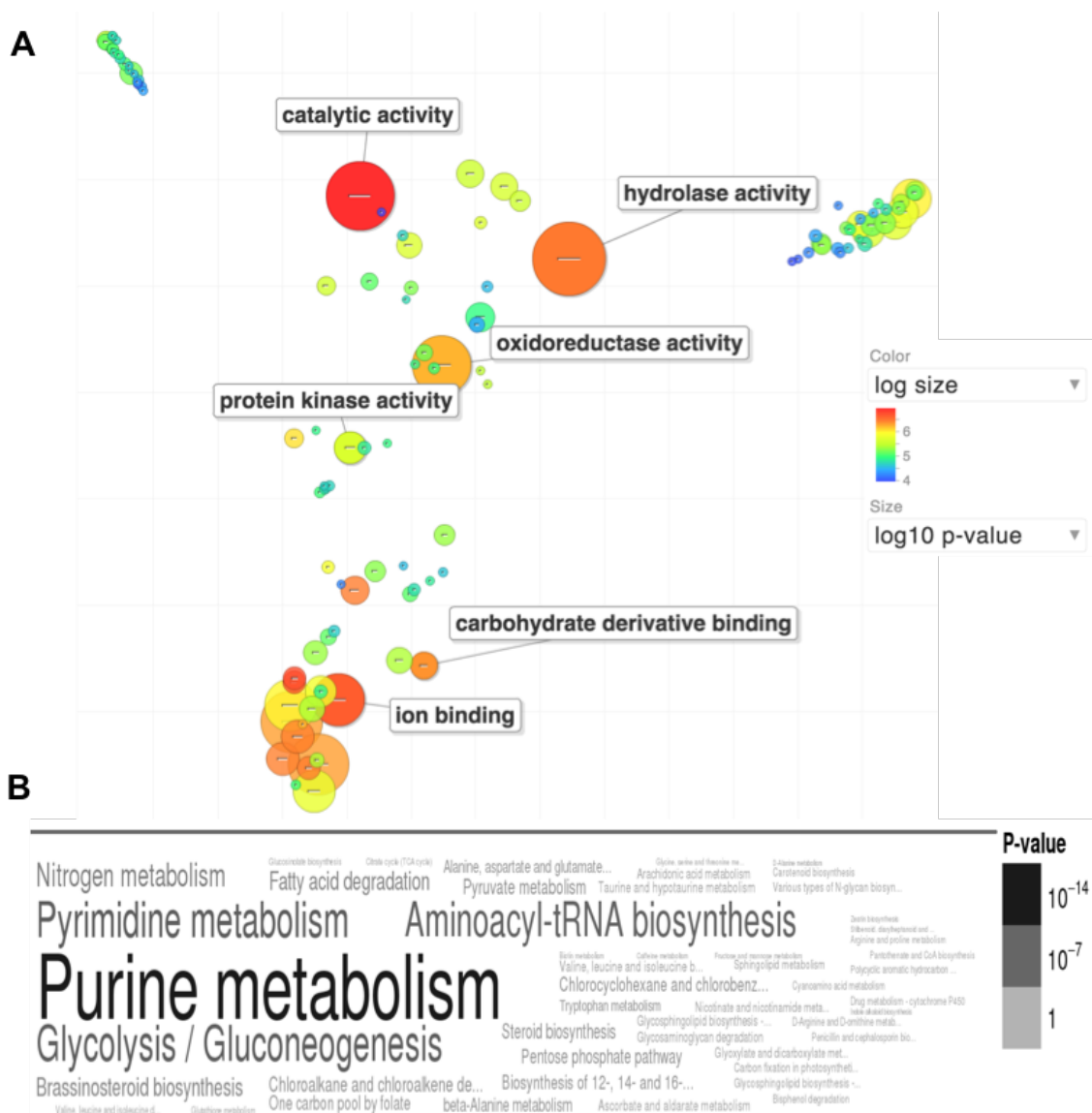
Figure 13: Gene ontology enrichment analysis of the 28 biomarker candidates indicates enriched hydrolase activity and increased purine and carbohydrate metabolism functions. A) Scatterplot of gene ontology enrichment of molecular functions, with size of the circles is proportional to the significance of enrichment. B) Word cloud of enriched KEGG metabolic pathways. Size of the words indicates significance of enrichment. Figures were produced using the Revigo[106] and GOSummaries[108] wordcloud plugin from FungiDB (www.fungidb.org).[107]

The 28 biomarker candidate proteins were then identified from *in vitro* fungal

cultures produced in both the mycelial and spherule forms, in three different media

combinations (Table 2). The mean NSAF of triplicate technical replicates for all

biological replicates (n), are provided. Also included in this table is information regarding

protein molecular weight, protein glycosylation and secretion signal predictions. Of the

28 biomarker candidates, 14 (50%) were predicted to contain N-glycosylation sites, 17

(60.7%) were predicted to contain O-glycosylation sites and 5 (17.9%) were predicted to

contain signal peptides. All 28 biomarker candidates are less than 65kDa in size, with 24

(85.7%) at being <45kDa in size, of which 11 (39.3%) are ≤20kDa in size.

Protein abundances in vitro and in vivo

At some abundance level, all 28 (100%) biomarker candidates were able to be

produced *in vitro* (Table 2). However, 1 of the biomarker candidates (Uncharacterized

protein CIMG_09001) was not present in *in vitro* grown spherules in any of the 3 media

used. Likewise, 1 of the biomarker candidates (Cytochrome c oxidase polpeptide VI) was

not present in any of the *in vitro* grown mycelial cultures but was present in *in vitro*

grown spherules. With regards to the in vitro growth of spherules, the media that

produced the most biomarker candidate proteins 27/28 (96.4%) was RPMI+ Survanta,

with Converse+Tamol producing 24/28 (85.7%) and RPMI+FBS produced 21/28

62

(75.0%) of thebiomarker candidates. With regards to the mycelial form, the media that produced the most biomarker candidate proteins was RPMI+FBS, which produced 25/28 (89.3%) of the proteins, whereas Converse+Tamol and RPMI + Survanta only produced 19/28 (67.9%) of the proteins in the mycelial form.

Table 2. Mean normalized spectral abundances (NSAF) and protein characteristics of the 28 biomarker candidates identified in this study. NSAF values provided are the mean of triplicate technical runs for all biolo[gical] replicates indicated (n).

#	Protein Name	Signal Peptide (Y/N)	N-glycan site?	O-glycan site?	Gene names	Molecular Weight	Mean NSAF of Triplicate Technical Replicates								
							In Vivo Spherules (LCM)			*In Vitro* Spherules			*In Vitro* Mycelia		
							Human (n=3)	Mouse (n=1)	Dog (n=1)	Converse+ Tamol	RPMI+FBS	RPMI+ Survanta	Converse+ Tamol	RPMI+FBS	RPMI+ Survanta
1	Hsp20/alpha crystallin family protein	N	N	Y	CPC735_047390	24 kDa	0.027128889	0.0052246	0.043804433	0.005164367	0.0033696	0.0067964	0	0.001582433	9.58333E-05
2	Peroxisomal membrane protein, putative	N	N	N	CPC735_024660	18 kDa	0.029539556	0.015784567	0.00329843	0.000588617	0.0040772	0.004162767	0.007332867	0.004248167	0.0028017
3	Coflin	N	N	N	CISG_03898	17 kDa	0.003289478	0.010636633	0.004105	0.005141667	0.003360567	0.00262067	0.000776703	0.001678467	0.0009374
4	Pyruvate decarboxylase	N	N	Y	CPSG_03493	63 kDa	0.003388875	0.0055436	0.0054683	0.006482167	0.011000733	0.0078734	0.000950207	0.003366567	0.002466967
5	Ag2/Pra CRoW domain	N	N	N	A1KXE9	8 kDa	0	0.0055391	0.008156	0.000478067	0	6.51033E-05	0.0049153	0.0037228	0.015192333
6	Uncharacterized protein CISG_02340	N	N	Y	CISG_02340	12 kDa	0.001977533	0.006869967	0	0	0	0.000950207	0	0.00042959	0
7	Uncharacterized protein CIMG_05576	N	N	Y	CPSG_05795	57 kDa	0.004773722	0.007456667	0	0	0	0.000105227	0	0	0
8	Uncharacterized protein CIMG_09001	Y	N	N	CPAG_07918	13 kDa	0.001772956	0.004655333	0	0	0.00060366	0	0.000444273	7.67733E-05	0
9	Uncharacterized protein CPC735_057740	N	N	N	CPC735_057740	16 kDa	0.0022172	0.0032966	0	0.000942047	0	0.00064826	0.00010131	0.00010131	0.000137557
10	Glycine-rich protein	N	N	N	CPC735_026500	36 kDa	0.001628233	0.0037654	0.0044632	0.003578067	0	0.0003831687	0	0.00044799	0.000243787
11	Uncharacterized protein CPC735_030710	N	Y	Y	CPC735_030710	21 kDa	0.001071156	0.0014889	0.00841813	0.002506633	0.0065222	0.0017747	0.000522653	0.00017747	0.007248333
12	ATP synthase subunit d, mitochondrial	N	N	Y	CPC735_012590	20 kDa	0.0010745444	0.004520267	0.00313977	0.001355077	0.00019474	0.003989387	0.000342353	0.000342353	0.000243787
13	Keto-acid reductoisomerase, mitochondrial	N	N	Y	CPC735_067570	45 kDa	0.002796653	0.003328533	0.00161113	0.007236333	0.004502833	0.005556733	0.001595467	0.0036316	0.000697267
14	Cytochrome c oxidase polypeptide VI	N	Y	Y	CPSG_00594	18 kDa	0.002470394	0.003292767	0.0040629	0.00011077	0	4.12867E-05	0	0	0
15	Fructose 1,6-bisphosphate aldolase, putative	N	N	N	CPC735_006240	40 kDa	0.003073342	0.00116935	0.00985197	0.005408767	0.002541433	0.0068675	0.001158173	0.0033835	0.000786367
16	NADP-dependent leukotriene B4 12-hydroxydehydrogenase	N	N	N	CPAG_08737	38 kDa	0.001009167	0.0022601	0.003356677	0.004155633	0.002336467	0.003146367	0.000282447	0.000872503	5.81067E-05
17	ATP synthase delta chain, mitochondrial,	N	N	N	CPC735_070430	18 kDa	0.0022601	0.002215467	0.0018784	0.000729167	0	0.00048104	0.000309333	9.02667E-05	0.00044799
18	Uncharacterized protein CISG_09979	N	Y	N	CISG_09979	13 kDa	0.003655125	0.0020165	0.005643733	0.0054964	0.004454233	0.004614533	0.000309333	0.001532333	0.000177377
19	Aha1 domain-containing protein	N	Y	Y	CPSG_01619	36 kDa	0.001016811	0.004302133	0.0006463	0.001250667	0.001241657	0.002360033	0.000170383	0.000503077	0
20	Protein wos2, putative	N	Y	Y	CPC735_022920	21 kDa	0.001740822	0.001913867	0	0.000062992	0.00088561	0.000834067	0.0000638973	0.000638973	0.000238317
21	Urease accessory protein ureG	N	Y	Y	CISG_10193	28 kDa	0.001325244	0.0030618	0	0.001113687	0.000127867	0.000941993	0.000170383	0.000638973	0.000238317
22	Profilin	N	N	N	CPC735_064080	14 kDa	0.000505278	0.005056667	0	0.004609267	0.004354767	0.003829433	0.003125367	0.001549033	0.00018926
23	Uncharacterized protein CIMG_00509	N	N	N	CPC735_057210	11 kDa	0.001643427	0.002322033	0	0	0	6.83933E-05	0.001458433	0	0.001370433
24	Dihydrodipicolinate synthetase	N	N	N	CPSG_04876	33 kDa	0.001229522	0.002786633	0	0.0001735563	0.00086479	0.000129088	0.00010364	0.000416043	0.000193193
25	Protein disulfide-isomerase	Y	Y	Y	CPSG_035440	57 kDa	0.002338111	0.0033413	0.00305353	0.00113174	0.001266557	0.001266557	6.74253E-05	0.001373713	3.86633E-05
26	Flavodoxin domain containing protein	N	Y	N	CPC735_070770	22 kDa	0.003039333	0	0	0.004272	0.001832867	0.00038938	0.0034799	0.002461433	0.002197833
27	putative	N	Y	Y	CPC735_074040	26 kDa	0.001260586	0.0036157	0.00049183	0.002193867	0.0034962	0.00218275	0.000592483	0.001465373	0.000334233
28	GTP-binding protein sarA, putative	Y	Y	Y	CPC735_069720	21 kDa	0.001788711	0.000629867	0	0.003123033	0.000358567	0.001047767	0.00053445	0.001089123	0

64

The mean NSAF abundances in Table 2 were used to calculate protein abundance fold changes between *in vivo* spherules and *in vitro* grown spherules and mycelia. In Figure 14A, a volcano plot shows fold change of protein abundance (NSAF) of the 28 proteins from *in vivo* spherules relative to the protein abundance from *in vitro* spherules. Five of the 28 biomarker candidate proteins were significantly more abundant in the *in vivo* spherules than *in vitro* grown spherules: Cofilin, Peroxisomal matrix protein, Cytochrome c oxidase polypeptide VI, Uncharacterized protein CISG_02340 and Uncharacterized protein CIMG_05576. The highest fold change was Uncharacterized protein CISG_02340 (~256 fold greater *in vivo* than in *in vitro* spherules).

Fold change of relative protein abundances from *in vitro* grown spherules in relation to *in vitro* grown mycelia is shown in Figure 14B. Eight proteins were significantly more abundant in the *in vitro* grown spherules than *in vitro* grown mycelia. The proteins with the greatest fold changes (~8 fold) include NADP-dependent leukotriene B4 12-hydroxydehydrogenase (LTB4D), Uncharacterized protein CISG_09979, Hsp20/alpha crystalin family protein, and Glycine rich protein. Ag2/PRA CRoW domain protein was ~7 fold higher in the mycelial form than the spherule form when grown *in vitro* (Fig.14B), but although not statistically significant, was ~12 fold higher in *in vivo* spherules than *in vitro* spherules (Fig.14A).

The most significantly abundant proteins from Figures 4A and 4B were then evaluated for their ability to be produced *in vitro*. Figure 15 (A-I) shows the protein

65
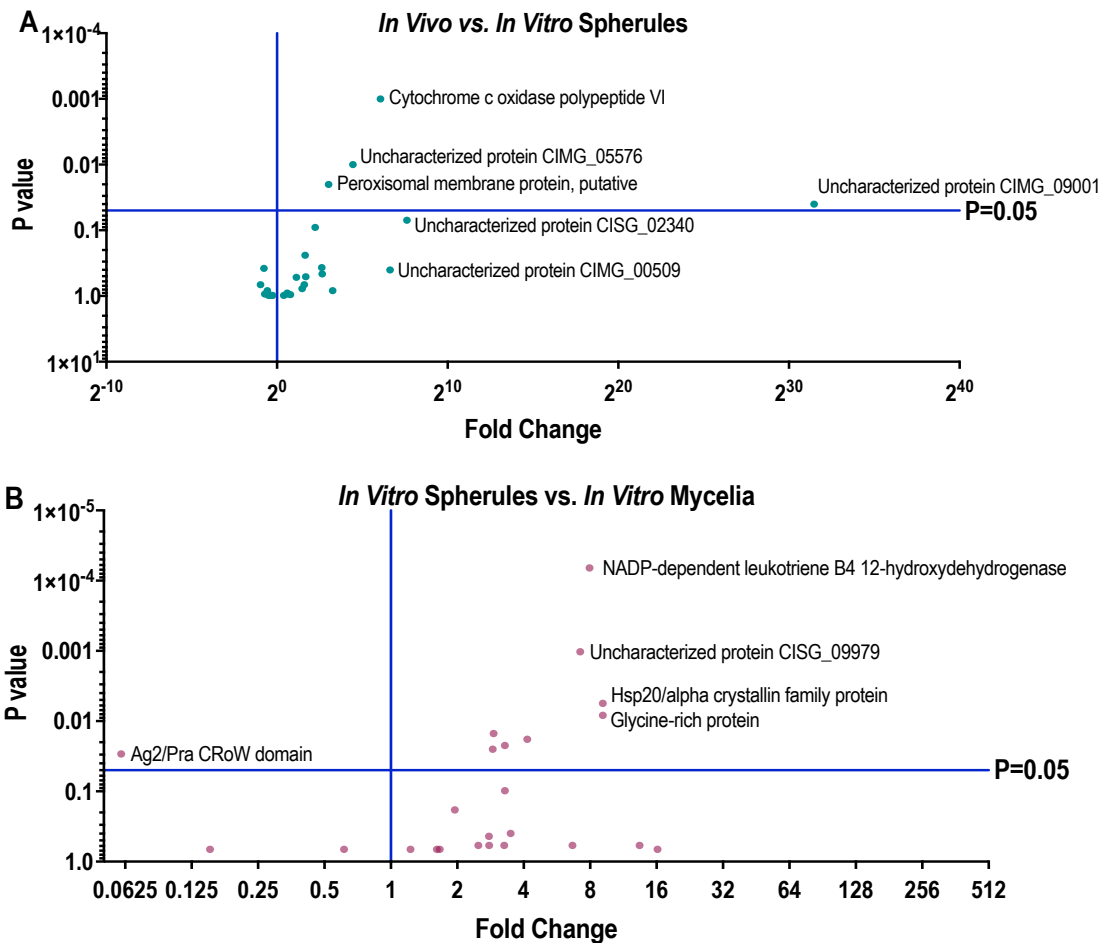
Figure 14. Volcano plots of the 28 biomarker candidate proteins indicates large fold change of protein abundances (NSAF) between *in vivo* and *in vitro* grown proteins. A) Fold change of relative protein abundances of LCM lung tissues in relation to the spherule form. B) Fold change of relative protein abundances in the spherule form in relation to the mycelial form. Volcano plots produced in GraphPad Prism v.8, using multiple t tests of NSAF, corrected for multiple comparisons using Holm-Sidak method.

abundances of each *in vitro* media used, as compared to the abundance of the same

protein *in vivo*. All of the 10 most significantly abundant *in vivo* spherule proteins were

able to be produced in spherules *in vitro*, however abundances were media dependent.
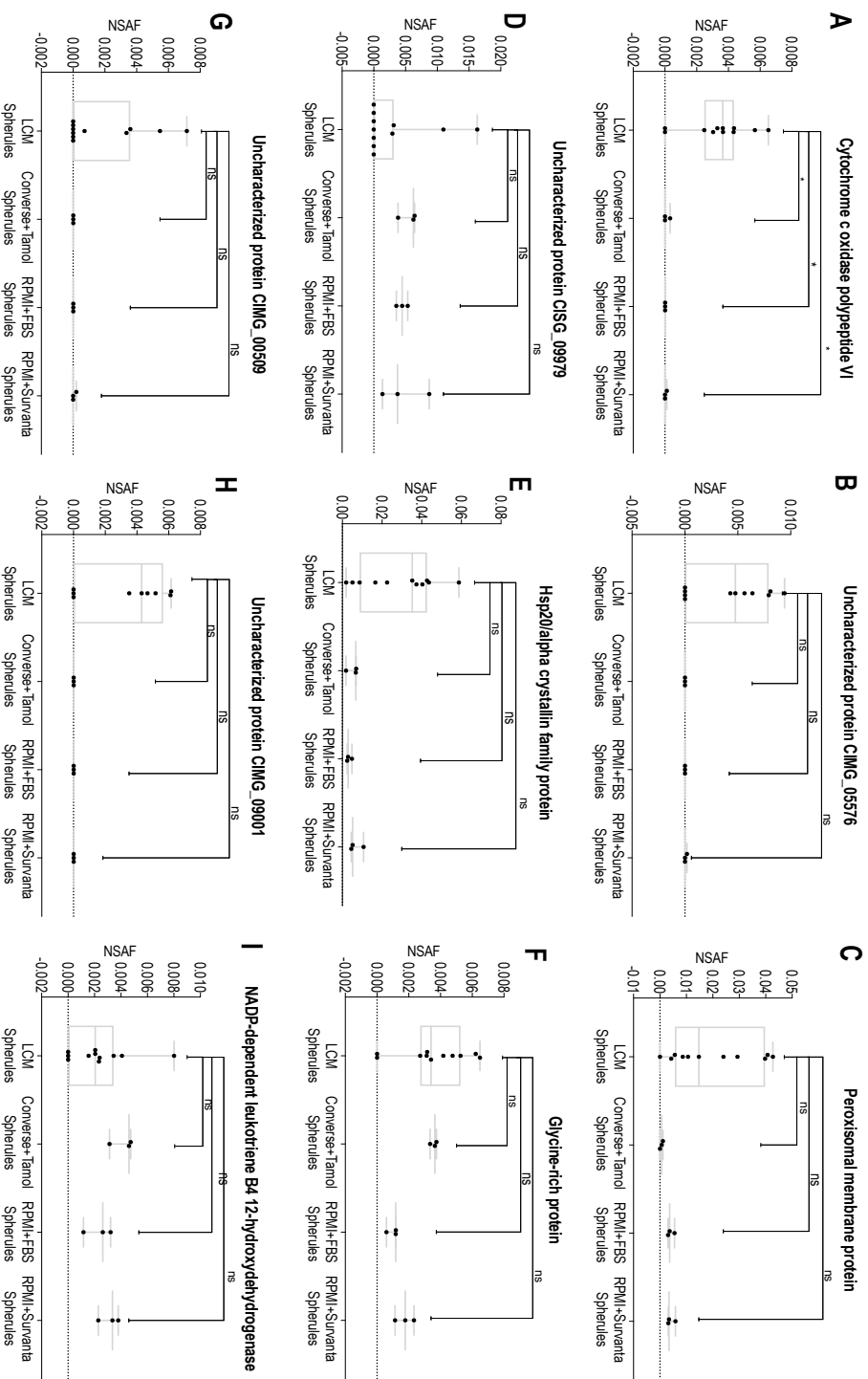
Figure 15: Box and whisker plots of differential protein abundances indicate that all biomarker candidate proteins can be produced *in vitro*, but abundance is dependent on culture media used. *In vivo* versus *in vitro* grown spherule abundances are shown for 8 proteins with the highest fold changes (A-I). Significance was calculated using one-way ANOVA adjusted for multiple comparisons using Sidak's method.

Discussion

In this chapter, the identification biomarker candidate proteins from *Coccidioides spp.*- infected mammalian lung tissues and the determination of whether these proteins are produced *in vitro* in any of the 3 growth media, was evaluated. By identifying the 100 most abundant proteins in LCM captured spherules and removing 72 human orthologues sharing ≥40% identity to human proteins, 28 possible *Coccidioides-spp.* biomarker proteins were identified. All 28 (100%) of the biomarker candidate proteins are *produced in vitro*, however, protein abundance was growth phase and media dependent. Additionally, three of the 28 biomarker candidates were found to be *Coccidioides-*specific biomarkers (Uncharacterized proteins CIMG_00509, CIMG_05576, CIMG_09001), as they do not share significant sequence identity to other fungal lung pathogens. Future studies which characterize the functions of these proteins is warranted.

To date, there has only been one other proteomic evaluation of *Coccidioides spp.* proteins from *in vivo* produced samples. In this study by Lewis *et al* (2015), who evaluated *Coccidoides spp.* proteins collected from bronchiolar lavage fluid (BALF) of laboratory-infected mice,[112] only 8 *Coccidioides spp.* proteins were identified. In fact, all 8 of the *Coccidioides spp.* proteins identified in BALF were also identified in our LCM lung tissue dataset. Remarkably, 2 of the 8 (25%) BALF proteins they identified (Peroxisomal membrane protein and Uncharacterized protein CIMG_09001) were in our 28 candidate biomarkers dataset. However the other 6 proteins they identified in BALF

either shared too much identity with human proteins or were not in the top 100 most abundant proteins from LCM lung tissues. It is interesting that Uncharacterized protein CIMG_09001, which we identified as a *Coccidioides*-specific biomarker candidate, was produced in BALF. This data gives credence to the LCM-assisted biomarker discovery methodology used in this chapter.

Barring the Lewis *et al* study, all other published proteomic studies of *Coccidioides spp*. have evaluated proteins produced from *in vitro* cultures, which are not truly representative of *in vivo* expression.[64,98,113 114] Additionally, all previous studies used only one culture medium; Converse media supplemented with Tamol. Interestingly, of the 28 biomarker candidates identified in this study to be abundantly produced *in vivo*, 4 (14.3%) were not produced in the spherule form and 9 (32.1%) were not produced in the mycelial form, when grown in Converse with Tamol *in vitro*. Previous studies which only used *in vitro* growth in Converse, are likely to have missed these proteins.

One of the aims of this study was to find a single growth medium which causes spherule protein expression to most closely resemble *in vivo* growth. However, none of the media tested closely mimics *in vivo* growth as spherules. Different *in vitro* culture media and conditions allowed for different subsets of *Coccidioides* proteins to be produced. Regardless, of the 3 media evaluated in this study, the maximum number of the 28 biomarker candidates was produced using RPMI+ Survanta (27/28; 96.4%) in spherule cultures.

RPMI media supplemented with Survanta has not been previously published for use with *Coccidioides spp. in vitro* cultures. Within this study it was used as an alternative surfactant to Tamol. Tamol is a synthetic anionic surfactant made from sulfonic acid salts, more commonly used as a paint dispersant. Tamol has been shown to induce faster spherulation in *in vitro* culture and cause reversion of mycelia into spherules.[115,116] Although the exact mechanism of Tamol-induced spherulation is not entirely known, it is likely due to its ability to emulsify media components and reduce surface tension around the spherule cell walls. Survanta is a natural bovine lung extract used in the prevention and treatment of respiratory distress syndrome in premature infants. In addition to containing surfactant proteins B and C (SP-B and SP-C), it contains phospholipids, neutral lipids, fatty acids, and surfactant-associated proteins which mimic the surface-tension lowering properties of natural lung surfactant. Additionally, as a working solution of 0.1% in RPMI contains only 0.1mg/ml protein, it is much more compatible with mass spectrometry than RPMI supplemented with 10% FBS.

In vivo grown spherules also produced protein abundances that differed from *in vitro* grown spherules. Cofilin, Peroxisomal matrix protein, Cytochrome c oxidase polypeptide VI, Uncharacterized protein CISG_02340 and Uncharacterized protein CIMG_005576 were all significantly more abundant *in vivo*. Cofilin activity has been shown to be required for pathogen entry into host cells, including *Cryptococcus neoformans* and *A. fumigatus* internalization into type II alveolar epithelial cells through

70

the RhoA-ROCK-LIM kinase pathway.[117] Peroxisomal matrix protein, in addition to

being found in mouse BALF, has also been investigated as a recombinant vaccine

candidate for VF.[98] Although it showed good protection against intraperitoneal challenge,

only modest protection was afforded following high dose intranasal challenge. However,

in addition to being biomarker candidates, any of the other 28 proteins have the potential

to also be vaccine candidates or targets for therapeutic drugs. Taken together, the findings

reported in this chapter suggest that Coccidioidal protein expression in vivo is distinct

from Coccidioidal protein expression in 3 different in vitro growth conditions.

Furthermore, my findings form a foundation upon which to select relevant biomarkers for

antigen-based detection of Coccidioides in potentially infected patients.

CHAPTER IV

CARBO-LOADING IN FUNGAL LUNG PATHOGENS: A QUANTITATIVE

ANALYSIS OF CAZYME ABUNDANCE AND RESULTING GLYCANS

Abstract

Coccidioidomycosis, also known as Valley Fever, is caused by two *Coccidioides*

species*: C. posadasii* and *C. immitis*. They are dimorphic fungi, existing in a mycelial

form in the soil and a yeast-like spherule form in mammalian hosts. *Coccidioides spp*. are

important and common pneumonia-causing pathogens of the American southwest, but

little is known about their glycobiology or how their glycosylations differ from other

pneumonia-causing fungi. There is mounting preliminary evidence to suggest genus or

even species-specific glycosylations in the fungal kingdom due to the presence of unique

carbohydrate active enzymes (CAZymes) in different fungal genomes.[67,118] If

Coccidioides spp.-specific glycans can be identified, it may be possible to exploit these

differences to develop more specific diagnostic approaches and more effective

therapeutics. Herein we i) identify *Coccidioides spp*.-specific CAZymes by

bioinformatically comparing the CAZyme repertoire (CAZome) of *Coccidioides spp*. to

other common fungal lung pathogens and a non-pathogenic close fungal relative, ii)

experimentally evaluate *Coccidioides spp*. CAZyme abundance *in vivo* and *in vitro*, and

iii) identify *Coccidioides* genus-specific N-glycans by experimentally determining the N-

glycan population (N-glycome) of *Coccidioides*-infected lung tissues using tandem mass spectrometry. As far as we are aware, this is the first use of mass spectrometry to compare the N-glycomes and CAZomes of different fungal genera during infection in human hosts.

Introduction

CAZymes play a crucial role in the formation and remodeling of complex carbohydrates within cells. N-glycans are carbohydrates attached to asparagine residues by a glycosidic bond and are the major constituents of glycoproteins in eukaryotes.[32] Owing to their presence on cell walls and secreted proteins, they are conveniently located at the site of cell-to-cell contact, and therefore play important biological roles in cellular communication (signaling), immune recognition/modulation and enzyme activity.[119] Approximately 3-4% of any given fungal genome codes for CAZymes, approximately twice the percentage that is in the human genome, highlighting the importance of sugar structures in function and viability in fungi.[120]

N-glycan core structures tend to be conserved, but considerable diversity exists in the elaborate branching of the oligosaccharides, even between relatively similar species.[121] Interspecies glycan diversity is thought to be involved in morphological and phenotypic trait differences, or reflect selective pressures resulting from evolution in different ecological niches. A recent mass spectrometry-based N-glycome study

comparing three *Aspergillus spp.* found interspecies variations with novel glycan branching.[66] Another study comparing various filamentous fungi was able to suggest species-specific variations in N-glycan synthesis due to absence or presence of certain mannosyltransferases.[67] To date, no genus- or species-specific glycosylations are known in dimorphic fungi, including *Coccidioides spp.*

      Coccidioides spp. cell walls are known to contain 1,3 β-D-glucan, chitin, and galactomannan. However, these glycans are also found in other fungi. Fungal diagnostic platforms that utilize these conserved oligosaccharides suffer from poor specificity, as they are not specific for a particular fungal genus.[30] Since infection with *Coccidioides spp.* causes ~30% of pneumonias in endemic regions, it is important to develop more specific diagnostic approaches and identify new targets for treatment. Herein, we explore the CAZome and the resultant N-glycome produced by *Coccidioides spp.,* in order to better understand metabolic processes that can be exploited for diagnostic or treatment purposes, and draw comparisons between other pneumonia-causing fungal pathogens or *Uncinocarpus reesii*, a non-pathogenic fungus that is the closest known genomic relative to *Coccidioides* spp.

Materials and methods

      An overview of the proteomics and glycomics workflows used in this study can be seen in Figure 16.

74
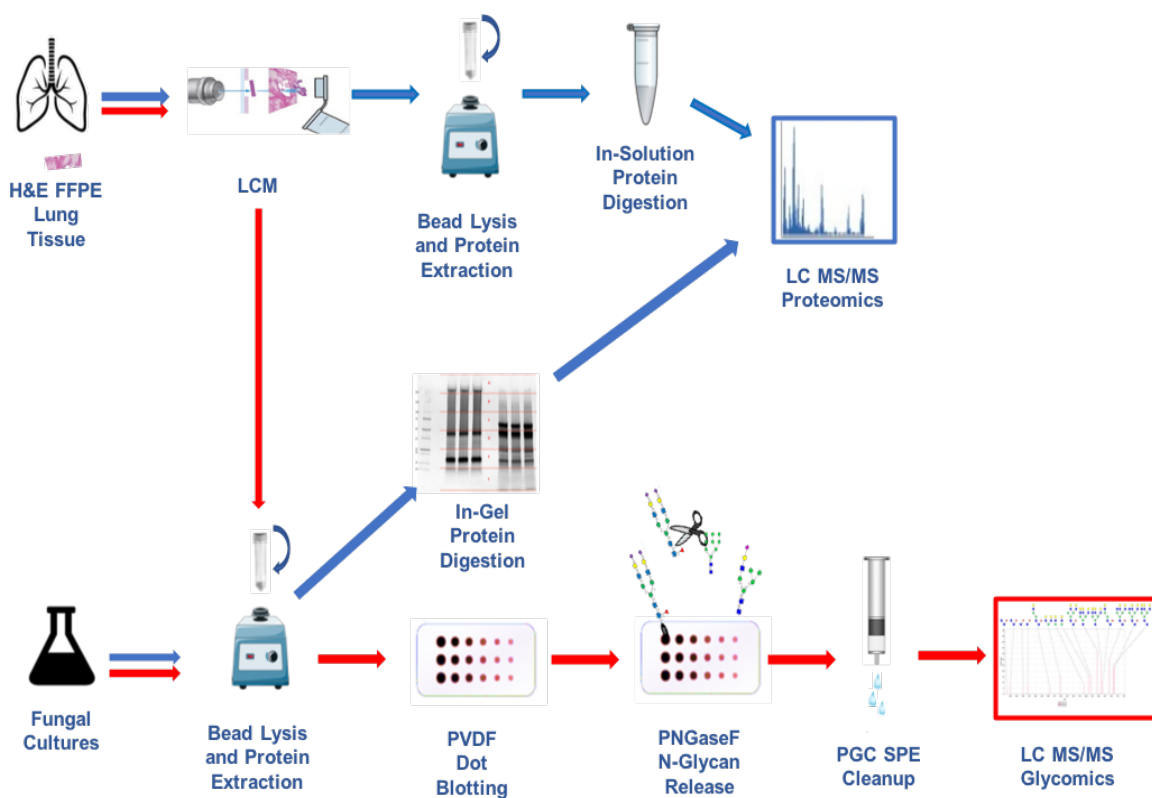
Figure 16. Schematic of proteomics and glycomics methods used in this manuscript. Triplicate technical runs of formalin fixed paraffin-embedded (FFPE) lung tissues from each of three patients infected with each of the following fungi (*Coccidioides spp.*, *Histoplasma spp.*, *Aspergillus spp.*, and *Cryptococcus spp.*) and three uninfected lung controls, were cut out by laser capture microdissection (LCM). Proteins were in-solution trypsin digested after bead lysis and analyzed by LC-MS/MS for proteomics (blue pathway), or blotted onto polyvinylidene fluoride (PVDF) membranes and deglycosylated with PNGaseF with subsequent desalting with porous graphitic carbon (PGC) solid phase extraction (SPE) for glycomics (red pathway). Mycelial and spherule form *in vitro* cultures of *Coccidioides posadasii* strain Silveira grown in different liquid culture media were used as positive controls. Similarly to LCM lung tissues, in-vitro fungal cultures were subjected to bead lysis and glycomics (red pathway). Lysates from fungal cultures were digested in-gel prior to LC-MS/MS proteomics (blue pathway).

FFPE lung samples

Formalin-fixed paraffin embedded (FFPE) human lung tissue blocks were acquired from Mayo Clinic Arizona histology biobank in accordance with IRB# 12-000965. Blocks were chosen from culture-confirmed clinical cases of fungal infections, or in the case of non-infected lung tissues, from autopsy samples of a non-infectious etiology. Three patients infected with each of the following fungi (*Coccidioides spp.*, *Histoplasma spp.*, *Aspergillus spp.*, and *Cryptococcus spp*.) and three uninfected patient controls were chosen, based on tissue quality and fungal burden. 10 μm thick sections were cut using a microtome and transferred to 1mm polyethylene naphthalate (PEN) membrane slides (Carl Zeiss Microscopy; Jena, Germany), deparaffinized and stained with hematoxylin and eosin stain (H&E) using standard procedures. Three technical replicates were performed for each of the 3 patients in each sample group; giving a total of 3 biological and 3 technical replicates for each sample used in this study.

LCM of FFPE tissues for glycomics

Laser capture microdissection (LCM) was performed using a Zeiss PALM MicroBeam scope with RoboPalm software (Carl Zeiss Microscopy). Approximately 500,000 μm^2 area of tissue was collected for each sample by laser capture. Tissue features were selected, collected and catapulted into the cap of 0.5 ml Eppendorf tubes containing 35μl of 0.1 M Tris-HCl (pH = 8.0) via laser pressure catapulting. Immediately

after capture, the tube was centrifuged at 14000 g for 2 min. to collect the lysis solution and tissue into the bottom of the tube and was frozen at -80C until processing.

Samples were processed by adding 35μl of 8% SDS, 0.2M DTT, 0.1M Tris-HCL (pH 8.0) lysis solution, to give a final concentration of 4% SDS, 0.1M DTT, 0.1M Tris-HCL (pH 8.0). Approximately 50μl of sterile 0.5m glass beads were added and samples were vortexed at maximum speed for eight cycles of 2min., with intermittent flash freezing on dry ice to break fungal cell walls. Formalin-protein crosslinks were broken from the tissue fragments by heating the sample at 99°C for 1 h in a thermalcycler.

"Dot Blot" N-glycan deglycosylation of LCM tissue

Methods were performed in accordance with the protocol published by Jensen, *et. al.* 2012,[105] with little deviation. In brief, glycoproteins from processed LCM FFPE tissues were spotted onto ethanol-wetted Immun-Blot® PVDF membrane (BioRad Laboratories; Hercules CA) and stained with Direct Blue 71 (MilliporeSigma; Burlington MA) to visualize proteins. After washing in 40% ethanol, 10% acetic acid, followed by water to remove acid, membranes were blocked in 1% (w/v) polyvinylpyrrolidone (PVP40; MilliporeSigma) in 50% (v/v)methanol. Protein spots were excised with a clean disposable scalpel and put into individual wells of a flat-bottom polypropylene 96-well plate (Corning; Corning NY). N-glycans were released from the protein spots by adding 5ul of PNGase F (New England BioLabs, Ipswich MA), 10μl of water and incubating in

a humidity chamber at 37°C overnight. The next day, the plate was sonicated in a water bath sonicator for 5 minutes prior to sample collection into microcentrifuge tubes. PVDF spots were washed twice with 20μl of water and washes were collected into the respective sample tube. Glycosylamines from the reducing terminus of PNGase F-released glycans were removed by the addition of 10μl of 100mM ammonium acetate (pH 5, MilliporeSigma) and incubation at RT for 1h. After drying samples in a SpeedVac concentrator at room temperature, free N-glycans were not reduced, but were desalted by using porous graphitic carbon (PGC) solid phase extraction (SPE) spin columns (Thermo Fisher Scientific, Waltham MA), using a modification to the Jansen *et. al.* protocol. Prior to sample loading, PGC SPE were washed in one column volume of 1M NaOH (MilliporeSigma), HPLC-grade water (Honeywell, Morris Plains, NJ), 30% acetic acid (MilliporeSigma), HPLC-grade water, and then equilibrated in 50% acetonitrile (ACN, Honeywell) with 0.1% triflouroacetic acid (TFA, MilliporeSigma) and then 5% ACN with 0.1% TFA. Samples were loaded onto the PGC SPE, spun out and then reloaded onto the column twice, prior to washing with one column volume of HPLC-grade water and one volume of 5% ACN with 0.1% TFA. Glycans were eluted by adding 1 column volume of 50% ACN with 0.1% TFA. Eluted glycans were dried in a SpeedVac concentrator at 37°C and frozen at -80°C until mass spectrometry.

LCM of FFPE tissues for proteomics

LCM was performed in the same manner as for glycomics, with the following modifications. Protein lysis solution contained 0.002% Zwittergent Z3-16 (MilliporeSigma) and lacked DTT. Instead, proteins were reduced by the addition of 1.8μl of 200mM tris(2-carboxyethyl)phosphine (TCEP, MilliporeSigma) in 0.1mM Tris, pH 8.2, (final conc. TCEP = 10mM) and incubated at 56°C . Proteins were alkylated by the addition of 1.9 μl of 200mM iodoacetamide (IAA, MilliporeSigma) in 0.1 M Tris-HCl (final conc. = 10 mM), and incubated in the dark at RT for 30 min. Proteins were digested by adding 2 μl of 0.01 μg/ul trypsin and incubating at 37°C for 16 h. The digestion was terminated by adding 4 μl of 2% TFA prior to loading on the mass spectrometer. Loading was normalized to the area of tissue, optimally 500,000 square microns per sample (~500 ng protein).

Fungal culture preparation and protein lysis

The mycelial form of *Coccidioides posadasii* strain Silveira (strain kindly provided by Dr. Bridget Barker of the University of Arizona) was grown in a shake flask rotating at 180RPM in 30°C for 7 days in one of the following six liquid media: 1) Levine modification of Converse media[104] with 0.5% N-Tamol (Dow Chemical Company, Midland MI), 2) modified Converse media with 0.1% Survanta (Beractant bovine lung surfactants; Abbot Laboratories, Chicago IL), 3) RPMI-1640 medium

(Corning) with 10% fetal bovine serum (FBS), 4) RPMI-1640 medium with 0.1% Survanta, 5) RPMI-1640 medium without supplementation, 6) modified Converse medium without supplementation.

Spherule form *C. posadasii* strain Silveira cultures were grown in vented shake flasks rotating at 180RPM in 20% CO_2 at 40°C for 7 days in the same six media as above. Cultures were centrifuged at 4,500RPM for 30min. in a Thermo Forma Multi RF tabletop centrifuge (Thermo) Supernatants were sterilized by filtration through 0.45μm vacuum filtration units (MilliporeSigma). Culture pellets were resuspended 5:1 in fungal lysis buffer (50 mM Tris, pH 7.6, 100 mM NaCl, 50 mM EDTA, 5% SDS) prior to transfer into 1.5ml silicone washer-sealed internally-threaded cryovials containing ~350μl of 0.5mm acid washed sterile glass beads. Cryovials were subjected to two rounds of 15 min. bead beating at maximum speed on a vortex with 1.5ml tube attachment unit followed by 3 cycles of flash freezing on dry ice. Sterility was checked by plating 10% of the total volumes onto 2x glucose yeast extract agar (2×GYE) for 3 weeks at RT.

Fungal culture protein extraction and in-gel protein digestion

Fungal culture proteins from filtered supernatants and pellet lysates were extracted using a modification of a previously published protocol.[105] In brief, filtrates and lysates were centrifuged for 30min. at 8,000RPM at 4°C. The supernatants were put into fresh tubes and proteins were precipitated by the addition of 4 volumes of ice cold

10%w/v trichloroacetic acid (TCA, MilliporeSigma) in acetone with 0.007% w/v

dithiothreitol (DTT, G Biosciences, St. Louis, MO). Samples were centrifuged at 3,000×g

for 10min. and the resultant protein pellets were washed three times in ice cold acetone

with 0.007% DTT. The final pellet was then resuspended in rehydration buffer [7M urea,

2M Thiourea, 4% w/v 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

hydrate (CHAPS, all MilliporeSigma) and 20mM DTT].

The samples in rehydration buffer were then filter exchanged with phosphate

buffered saline (PBS), using Amicon® Ultra 0.5ml 3kDa filtration units

(MilliporeSigma) prior to protein estimation using a micro bicinchoninic acid (BCA)

protein assay kit (Thermo). ~10μg of protein from filtrate samples and ~30μg of

corresponding lysate protein were mixed with reducing Laemmli sample buffer (BioRad,

Hercules, CA), heated at 95°C for 5min. prior to loading on mini-Protean™ TGX precast

gels(BioRad). Following electrophoresis, gels were stained with Bio-Safe Coomassie G-

250 Stain (BioRad) as per manufacturer's instructions. Each sample lane of the SDS-

PAGE gel was cut into six equal size fractions. Band fractions were then further reduced

into cubes of 1–2mm^3, put into low protein binding tubes (Eppendorf, Hamburg,

Germany) prior to destaining. Proteins were reduced in 10mM DTT for 30 min. at 60°C,

and alkylated with 55mM ioadacetamide (IAA, MilliporeSigma) for 30min. at room

temperature in the dark, prior to 37°C overnight digestion with Pierce™ MS Grade

trypsin protease (Thermo) diluted to 20ng/ml in 100mM ammonium bicarbonate

(MilliporeSigma). The peptides were then extracted from the gel pieces using 5% formic acid (FA), ammonium bicarbonate and acetonitrile washes prior to being dried down in a speed vacuum and stored at -80°C until LC-MS analysis.

Proteomic analysis of fungal lysates and filtrates

Protein digests were reconstituted in 0.1% FA and analyzed using LC-MS/MS by loading onto a Dionex UltiMate® 3000 RSLC liquid chromatography (LC) system (Thermo, San Jose, CA) using a PepMap RSLC C18 2um, 75um x 50cm EASY-Spray™ column (Thermo). Peptides were separated using a 0.3 μL/min LC gradient comprised of 2%-90% mobile phase B in 0-120 min. Mobile phase A was 0.1% FA in water and mobile phase B was 0.1% FA in acetonitrile. Eluting peptides were directly injected into an Orbitrap Elite Velos mass spectrometer (Thermo) and ionized using collision-induced dissociation (CID) of helium in positive ion mode. A "top 15" data-dependent MS/MS analysis was performed (acquisition of a full scan spectrum followed by collision-induced dissociation mass spectra of the 15 most abundant ions in the survey scan.

Protein Identification and label-free protein quantification

Database searching was performed using Sequest (Thermo) in Proteome Discoverer v1.4.1.14 (Thermo) against a combined FASTA database of all the most recent Uniprot *Coccidioides spp*. proteomes (*Coccidioides immitis* RS, proteome ID:

82

UP000001261, June 12, 2018 release date; *Coccidioides immitis* RMSCC 3703, proteome ID:UP000054559, February 26, 2018 release date; *Coccidioides immitis* H538.4, proteome ID: UP000054563, February 26, 2018 release date; *Coccidioides immitis* RMSCC 2394, UP000054565, February 26, 2018 release date; *Coccidioides posadasii* strain RMSCC 757 / Silveira, proteome ID: UP000002497, February 26, 2018 release date, *Coccidioides posadasii* C735, proteome ID: UP000009084, November 9, 2018 release date; and *Coccidioides posadasii* RMSCC 3488, UP000054567, October 26, 2018 release date). Searches were performed using a fragment tolerance of 0.60 Da (Monoisotopic), parent tolerance of 10 ppm (Monoisotopic), with Carbamidomethyl of cysteine as a fixed and Oxidation of methionine as variable modifications and a maximum missed cleavages allowed of 2. Protein identifications were accepted if they achieved a minimum of 2 peptides per protein and a false discovery rate (FDR) of <1%. Label-free protein quantification was performed using normalization of spectral abundance factors (NSAF) in Scaffold (v4.8.7, Proteome Software Inc.). The provision of information relating to mass spectrometry is in accordance with "Minimum Information About a Proteomics Experiment" (MIAPE) guidelines [79] and the mass spectrometry glycomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[80] with the dataset identifier PXD012459.

Glycan LC-MS/MS

Dried glycan extracts from dot-blotted and deglycosylated LCM tissues were solubilized in 10mM NH_4HCO_3 and analyzed by capillary-flow liquid chromatography electrospray tandem mass spectrometry (capLC-ESI-MS/MS) using a Thermo Fisher Scientific Q-Exactive Plus Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to a Thermo Ultimate 3000 RSLCnano HPLC system configured with the capillary flow module. The glycan mixture is loaded onto a Hypercarb Porous Graphite Carbon column (Thermo) and chromatography is performed using 10mM ammonium bicarbonate buffer in both the A solvent (98%water/2%acetonitrile) and B solvent (80% acetonitrile/20% water),starting at 10%B and increasing to 20%B in 5 minutes then a 20%B to 55%B gradient over 60 minutes at 4.8 μl /min.

The Q-Exactive mass spectrometer experiment was a positive mode data dependent set up with an MS1 survey scan from 380-2000 m/z at resolution 70,000 (at 200m/z), AGC target of 1e6, and maximum injection time (MaxIT) of 75ms, followed by HCD MS/MS scans at resolution 17,500 on the top 13 ions having a charge state of +2, +3, or +4, with an AGC target of 2e5, and max IT of 300ms . This is followed by a second positive mode data dependent series of an MS1 survey scan from 600 -1600 m/z at 70,000 resolution, triggering HCD MS/MS scans on the top 5 ions having a charge state of +1 or +2. The ions selected for MS/MS are placed on an exclusion list for 20 seconds.

Glycan identification

Glycan MS/MS spectra were analyzed using SimGlycan v.5.60 (Premier Biosoft; Palo Alto, CA). Search parameters were underivatized, free glycans in positive ion mode with H+ and Na+ adducts. A 1 Da ± 0.5 Da error tolerance was allowed. All glycosidic and cross-ring options were set to "yes". Only precursor mass to charge ratios (m/z) which also had MS2 spectra were kept for analyses.

CAZyme identification and alignment

CAZymes were identified from fungal proteomes by downloading FASTA for all associated proteins of all available strains in Uniprot.[122] The proteomes evaluated were: 2 *Aspergillus fumigatus* proteomes (UP000053005, UP000028372), 6 *Aspergillus niger* proteomes (UP000009038, UP000006706, UP000068243, UP000253845, UP000197666, UP000236662), 4 *Blastomyces dermatitidis* proteomes (UP000002039, UP000053869, UP000007802, UP000002038), 4 *Coccidioides immitis* proteomes (UP000001261, UP000054565, UP000054563, UP000054559), 2 *Coccidioides posadasii* (UP000002497, UP000054567), 3 *Cryptococcus gattii* (UP000053392, UP000053939, UP000053019), 10 *Cryptococcus neoformans var. grubii,(* UP000198225, UP000199770, UP000198245, UP000199744, UP000199721, UP000199748, UP000198246, UP000199747, UP000199756, UP000232048), 4 *Histoplasma capsulatum* (UP000002624,

UP000008142, UP000009297, UP000001631), and 1 *Uncinocarpus reesii* proteomes (UP000002058).

Downloaded FASTA were then searched in dbCAN (HMMdb release 7.0), a meta server for automated carbohydrate-active enzyme annotation.[123,124] CAZymes from *Aspergillus fumigatus*, *Aspergillus niger*, *Blastomyces dermatitidis*, *Coccidioides immitis*, *Coccidioides posadasii*, *Cryptococcus gattii*, *Cryptococcus neoformans var. neoformans*, *Histoplasma capsulatum*, and *Uncinocarpus reesii* genomes were explored. Searches were performed using all 3 available tools: HMMER, DIAMOND and HOTPREP databases, using default E-value cutoffs. As previously published, the genomes of Coccidioides posadasii strains are incomplete and have numerous annotation errors.[125] Accessions identified as CAZymes in only *C.immitis* and not in *C. posadasii* were checked by BLASTp sequence alignment and re-entering the protein FASTA in dbCAN for CAZyme matches.[126] Thus CAZyme presence/ absence was corrected accordingly in the provided data. CAZymes found to be truly unique to or absent in the homologue searches of *Coccidioides spp*. were further investigated for function and homology to other fungi.

Statistical analyses

GraphPad Prism v.7.0d was used for production of icemaps, charts and for 2-way analysis of variance (ANOVA) statistics. Monosaccharide composition analysis was performed using a separate in-house built R program, available upon request.

Results

Genome-wide CAZyme comparisons

A total of 9,950 CAZymes were identified from all *Aspergillus fumigatus*, 7495 from all *Aspergillus niger*, 2415 from all *Blastomyces dermatitidis*, 959 from all *Coccidioides immitis*, 785 from all *Coccidioides posadasii*, 3305 from all *Cryptococcus gattii*, 3205 from all *Cryptococcus neoformans var. grubii*, 2108 from all *Histoplamsa capsulatum,* and 532 from all *Uncinocarpus reesii* available sequenced genomes. As seen in Figure 17a, these proteins included CAZymes from all five of the known classes (AA, auxiliary activities; GH, glycoside hydrolase; GT, glycosyltransferase; and PL, polysaccharide lyase), as well as additional carbohydrate binding modules (CBMs). No members of the PL family were identified in the genomes of *Coccidioides, Uncinocarpus*, *Blastomyces* or *Histoplasma spp*.
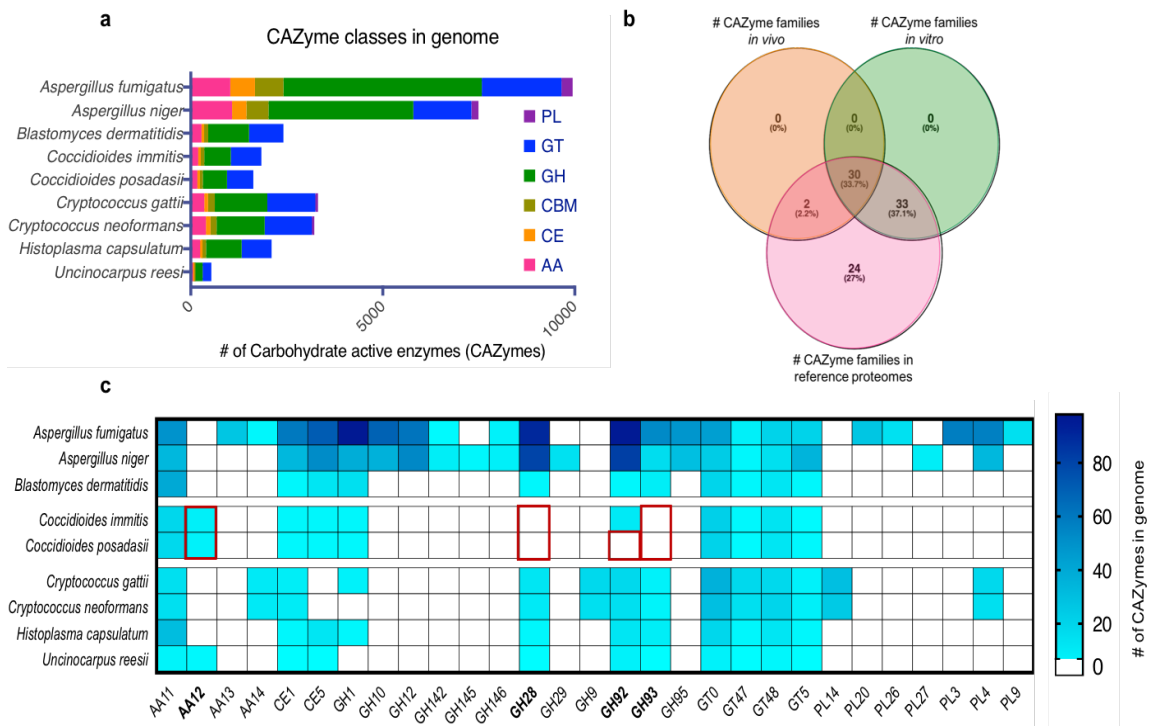
Figure 17. CAZyme classes and families encoded in the 7 sequenced genomes of *Coccidoides spp.* are largely expressed *in vitro* and *in vivo*, and indicate genus- and species-specific CAZyme profiles. (**a**) CAZyme classes identified in a genome-wide analysis, (**b**) CAZyme families found experimentally from *in vitro* grown *C. posadasii* strain Silveira or *in vivo* from *Coccidioides spp.*-infected lung tissues, expressed as a percentage of total CAZyme families found in the 7 *Coccidioides spp.* genomes, and (**c**) CAZyme family icemap showing the number of CAZymes in the genomes of all 7 published *Coccidioides spp.* sequences and closely related fungi. All fungi evaluated are able to cause pneumonia in humans, except *U. reesii,* a non-pathogenic genomic relative of *Coccidiodes spp.* *Coccidioides spp.*-unique CAZyme profiles are highlighted within red boxes. AA,auxiliary activities; GH, glycoside hydrolase; GT, glycosyltransferase; CBM, carbohydrates- binding module; PL, polysaccharide lyase.

Experimental evidence of CAZyme expression

The combined reference proteomes of all 7 sequenced *Coccidioides spp.* reference proteomes available in the Uniprot database were found to contain 89 CAZyme families (*ie* AA1, AA2, etc.). Proteins from 63 (70.8%) CAZyme families were confirmed experimentally by mass spectrometry from *Coccidioides spp* CAZyme proteins. Sixty-three (70.8%) were identified from *in vitro* grown *C. posadasii* strain Silveira lysate, 32 (36.0%) were identified in *Coccidioidies spp.*- infected human lung tissues, and 30 (33.7%) were in both *in vitro* and *in vivo* samples (Fig. 17b).

Identification of Coccidioides genus-specific CAZymes

As shown in Figure 17c, an icemap of select CAZyme families indicates that *Coccidioides spp.* genomes contain a distinctive combination of CAZymes, which differ from other fungal pathogens and closely related non-pathogenic *Uncinocarpus spp.* Uniquely, family AA12 is present and family GH 28 and GH93 are absent in all *Coccidioides spp.* Family AA12 is present in both *Coccidioides spp.* and *U. reesii*, but is absent in the other fungi examined, whereas families GH28 and GH93 are absent from *Coccidioides spp.*, but present in all the other fungi evaluated. Finally, CAZyme family GH92 was present in the genome of all fungi examined except *C. posadasii*.

In-vitro and in-vivo CAZyme abundance profiles

There was a combined total of 2,154 proteins identified by mass spectrometry analyses of *C. posadasii* strain Silveira cultured using 6 different media: 1) Converse with Tamol, 2) Converse with Survanta, 3) RPMI+10%FBS, 4) RPMI+Survanta, 5) RPMI alone, 6) Converse without Tamol. The 6 *in vitro* culture conditions of *C. posadasii* strain Silveira contained 111 CAZymes combined, and a further 47 CAZymes from 1,421 proteins were quantified in *Coccidioides spp.*-infected human lung tissues by LCM (laser capture microdissection)-LC-MS/MS. CAZyme abundance profiles from *in vitro*-grown mycelia and spherules or *in vivo*-grown human LCM lung tissues were quantified using normalized spectral abundance factors (NSAF), a label-free quantification method of protein abundance.[111] An icemap of CAZyme protein abundances of all *in vitro* and *in vivo* LCM samples can be seen in Figure 18a. CAZymes GH18, GH20, GH38, GH47 and GH55 had a significantly higher mean abundance in the mycelial form isolates than the mean of the spherule form isolates or infected LCM tissues. Conversely, the mean abundances of AA3, GH17 and GH36 were significantly higher in the spherule form than the mycelial form. The CAZyme abundance differences between mycelia and spherules determined to be statistically significant were explored for their identities and enzyme activities, and are provided in Figure 18b. A table of all CAZymes and protein abundances are provided as Appendix D.

90

Figure 18. CAZyme abundances in *Coccidiodes posadasii* strain Silveira are significantly different between mycelial and spherule forms and depend on culture media used. (a) Icemap of CAZyme abundance profiles using normalized spectral abundance factors (NSAF) from *in vitro* mycelia and spherules grown in 6 different culture media or *in vivo* LCM lung tissues. Culture media: 1) Converse with Tamol, 2) Converse with Survanta, 3) RPMI+10%FBS, 4) RPMI+Survanta, 5) RPMI alone, 6) Converse without Tamol. **(b)** Description of significantly different CAZyme abundances between spherule and mycelial forms, with those marked with # being significantly more abundant in spherule form. Statistical analyses were performed using 2-way ANOVA with Tukey's multiple comparisons test of in vitro grown mycelia versus in vitro grown *C. posadasii* strain Silveira spherules grown in different media versus and vivo laser captured Coccidioides-infected lung tissues. Significance was determined as ***P<0.0001, ** P<0.001, *P<0.05.

In vivo N-glycomes of fungal-infected and non-infected lung tissues

      Glycan m/z values that were present in at least two of three biological replicates of lung tissue samples, but were not in other fungal-infected groups were assessed. A comparison of the most prevalent glycan m/z values for all fungal-infected LCM lung tissue groups or normal lung LCM controls, is shown in Figure 19. Only glycan m/z values unique to each sample group are shown. Normal human lung tissues contained 664 unique glycan m/z values, whereas human lungs infected with other fungi contained fewer total glycans. Lung samples infected with *Histoplasma spp*. contained 357 unique glycan m/z values, *Coccidioides spp*. contained 298, *Cryptococcus spp*. contained 292, and *Aspergillus spp*.-infected lungs contained 242 unique glycan m/z values. Regardless of lung tissue sample group, only 33-43% of all unique glycan m/z values were known glycans identified in the glycan databases.
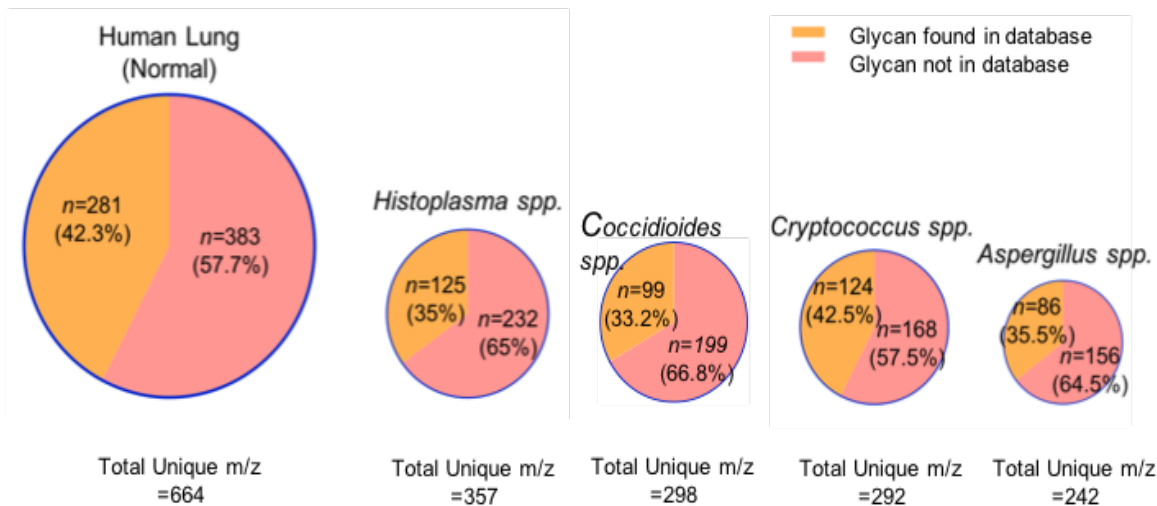
Figure 19. The majority of N-glycans in all fungal-infected and non-infected lung tissues are unknown glycan structures. Distribution of unique glycan m/z values found in at least 2 of 3 biological replicates, but not in other sample groups. Glycans which were identified in the SimGlycan database versus unknown glycan m/z values, by lung tissue sample, are shown. Size of pie is proportional to total number of unique m/z identified in the sample.

Monosaccharide composition analysis

For glycans identified in the glycan databases, monosaccharide compositions were calculated and normalized to the total number of glycans present in the sample (Fig. 20). A significant difference in N-acetylglucosamine (GlcNAc) abundance between mycelial ($p=0.0027$) and spherule ($p=0.0102$) lysates and normal human lungs was observed. Additionally, galactose was significantly more abundant in spherule lysates ($p=0.0224$) than in normal human lungs.
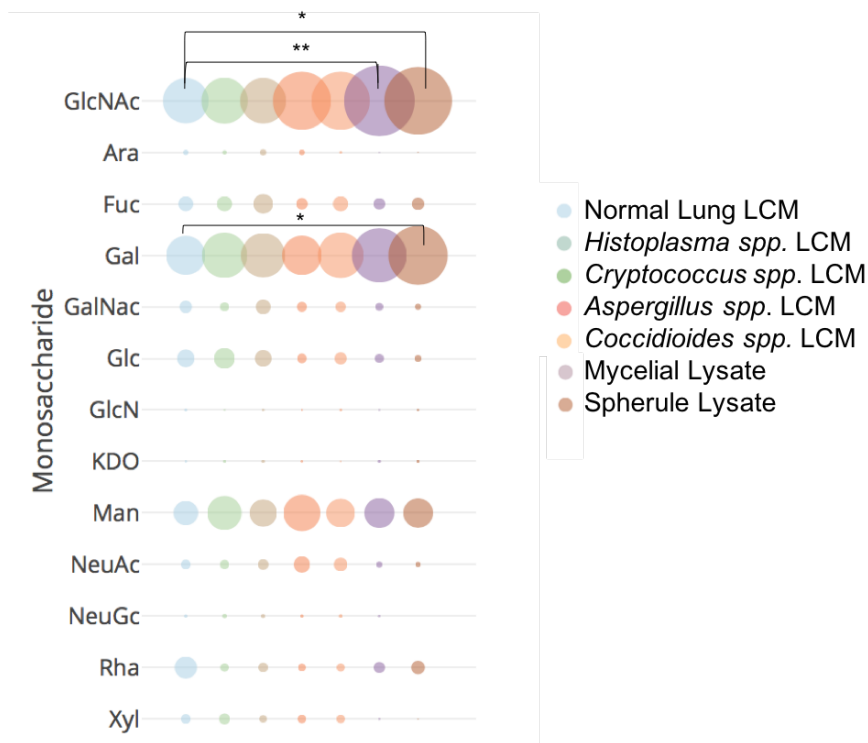
Figure 20. Galactose and N-acetylglucosamine monosaccharides are significantly more abundant in *Coccidioides spp.* than in human lung tissues. Monosaccharide compositions of glycans that matched the glycan database are presented. Monosaccharide values were normalized to the total number of glycans present in the sample. Statistical significance was determined by 2-way ANOVA, using Dunnett's multiple comparisons test against normal lung tissue. ***$P<0.0001$, ** $P<0.001$, *$P<0.05$. GlcNAc, n-acetylglucosamine; Ara, arabinose; Fuc, fucose; Gal, galactose; GalNac, n-acetylgalactosamine, Glc, glucose, GlcN, glucosamine; KDO, keto-deoxyoctulosonate; Man, mannose; NeuAc, n-acetylneuraminic acid; NeuGc, n-glycolylneuraminic acid; Rha, rhamnose, Xyl, xylose.

Glycan motif analysis

Glycan motifs of the known glycans were also evaluated for each of the infected

lung tissue sample groups (i.e. Aspergillus spp., *Coccidioides* spp., *Cryptococcus* spp., or

Histoplasma spp.) and control non-infected lung tissue. As shown in Figure 21a, there are

visible differences in the percentage of common glycan motifs between the different

fungal-infected lung tissues when compared to normal human lung. When comparing

only the glycan motifs of *Coccidioides spp*. samples to normal lung (Fig. 21b), a

significantly greater amount of the N-glycan Core Basic motif is found in the mycelial

(p=0.003) and spherule form lysates compared to normal uninfected human lung

(p=0.0002). Similarly, there is increased abundance of the N-glycan Complex motif in

both the mycelial *(p=0.0039)* and spherule form lysates *(p<0.0001)* compared to non-

infected lung.
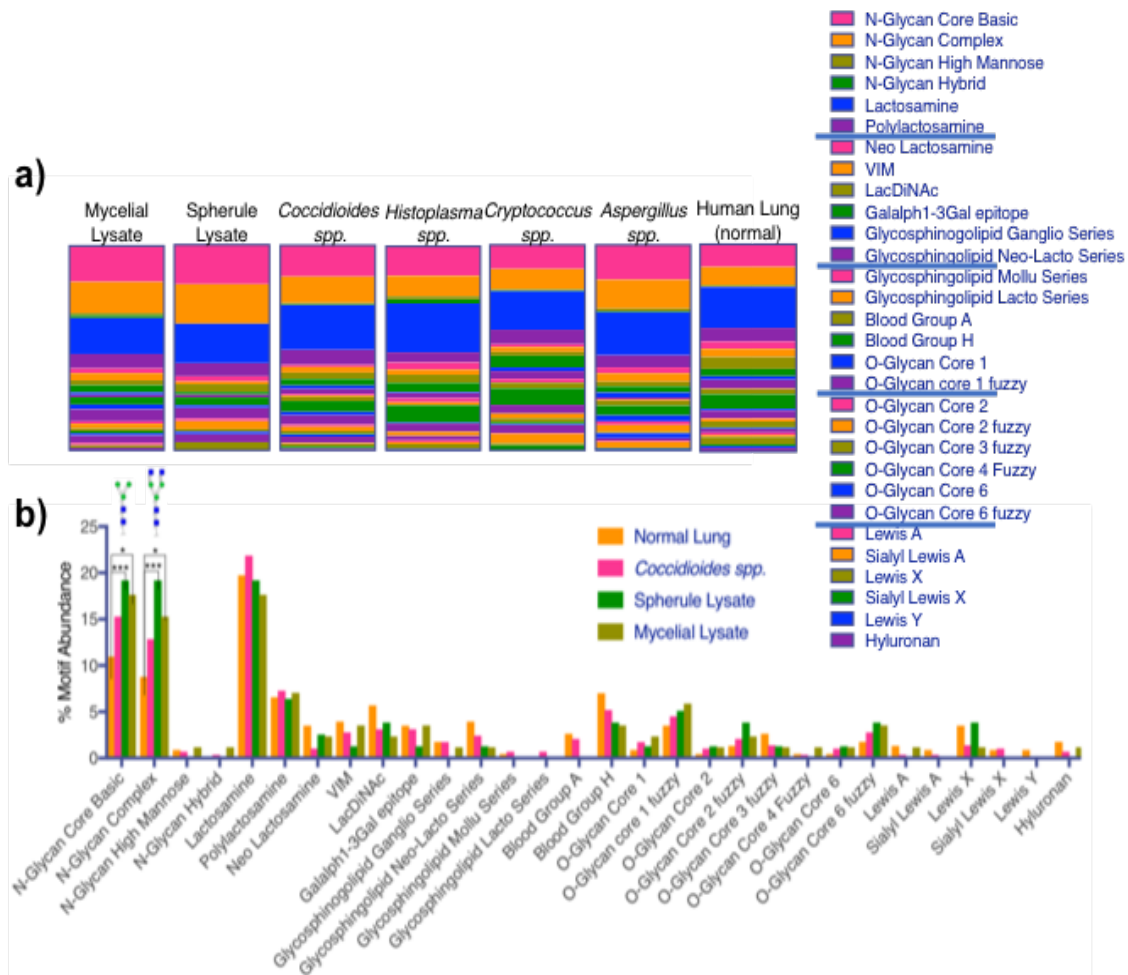
Figure 21. Glycan motifs of *Coccidioides spp.* from infected lung tissue differ from glycan motifs found in other fungi from infected lung tissue. Fungal-infected lung tissue and non-infected lung tissue have significantly different motifs. **(a)** Distribution of glycan motifs of known glycans found in each sample group as a percentage of total glycan motifs identified in each respective group. **(b)** Motif differences among non-infected lung LCM tissue, *Coccidioides spp.* LCM tissue and spherule and mycelial lysates of *Coccidioides posadasii* strain Silveira. Data excludes m/z values found within other sample groups or controls. Significance was determined by 2-way ANOVA. ***$P<0.0001$, ** $P<0.001$

Identification of Coccidioides genus-specific N-glycans

Of the 298 glycan m/z values unique to *Coccidioides spp*. samples, 14 were found in all 3 *Coccidioides*-infected patient lung tissues and *in vitro*-grown lysates (*p*=0.0003). Of these 14 *Coccidioides spp*.-unique glycan m/z values, 7 were not previously published glycans. The remaining 7 *Coccidioides spp*.-unique glcan m/z values correspond to 4 known glycans (3 of the m/z values matched the same glycan). These *Cocccidiodies spp*. glycans and resultant MS2 spectra are shown in Figure 22 a-d and represent potential markers that differentiate non-infected lung and lungs infected with other fungi.
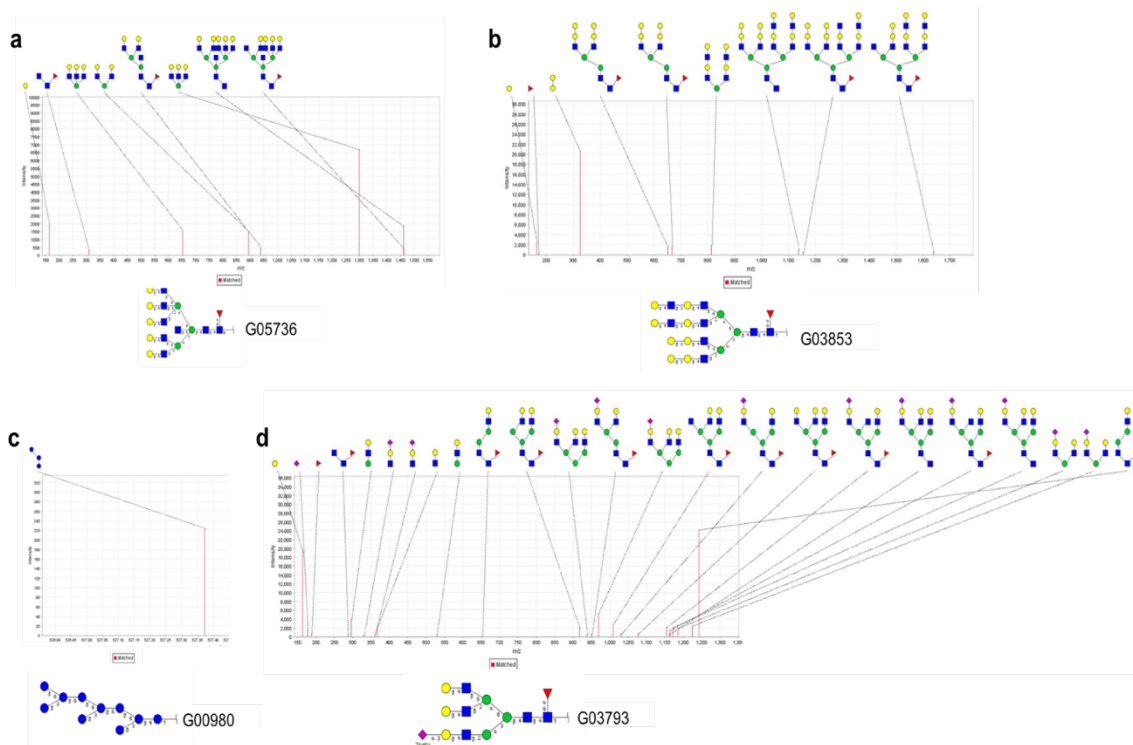
Figure 22. Mass spectrometry reveals *Coccidioides spp.*-specific glycan structures from infected lung tissues. (a-d) Four highly significant *Coccidioides spp.*-specific glycans ($p<0.001$) were identified in all *Coccidioides spp.*-infected lung tissues and *in vitro C. posadasii* lysates, but not in normal lungs or lungs infected with other fungi. Kegg Glycan identifiers and resultant MS2 spectra are provided for each glycan.

Discussion

In this study, we identified 89 CAZyme families encoded in the 7 available sequenced genomes of *Coccidioides spp.* When compared to other fungal pathogen genomes and the closest non-pathogenic genomic relative, *Uncinocarpus reesii*, *Coccidioides spp.* had a distinctive combination of CAZymes. The AA12 family CAZyme is found exclusively in all *C. immitis*, *C. posadasii* and *U. reesii* genomes,

98

corresponds to an L-sorbosone dehydrogenase (all homologues of CIHG_01000). This dehydrogenase is involved in the conversion of sorbitol and sorbose into ascorbic acid by converting L-sorbosone into 2-keto-L-gulonic acid.[127] Interestingly, L-sorbose degradation was among the transcriptionally enriched metabolic pathways among 152 up-regulated genes in a metanalysis study of two transcriptomics studies of spherule form *Coccidioides immitis*.[128] However, although the AA12 CAZyme was encoded in both the *C. immitis* and *C. posadasii* genomes, the protein was not identified *in vitro* or in LCM of *Coccidioides*-infected lung tissues. It is possible that translation of this protein is under post-transcriptional regulation and that under different culture media or growth conditions, this protein might be produced. Also, a lack of protein evidence by mass spectrometry does not preclude it from it having been produced in these conditions, but it is indicative that the abundance was low, if present at all.

CAZymes from CAZy families GH28 and GH93 were found to be encoded in the reference proteomes of all fungi evaluated in this study, except *Coccidioides spp*. Members of the GH93 family are listed in the CAZy database as exo-arabinases which hydrolyse linear α-1,5-L-arabinan frequently found in plant cell walls, but little other structural or mechanistic information is currently available.[129] The GH28 family of CAZymes include polygalacturonases (EC 3.2.1.15) and endo-rhamnogalacturonases (EC 3.2.1.171) involved in the degradation of plant biomass pectins.[130] Genes that code for the GH28 family of enzymes are highly prevalent in Ascomycota and Basidiomycota,

with higher copy numbers in saprotrophic fungi than plant and human pathogens, suggesting their important role in those organisms.[131] Although enzymes of the GH28 family catalyze pectin by hydrolysis and members of the pectin lyases (PL) family catalyze by ß-elimination, they are structurally similar.[132] No members of the PL class of CAZymes were found in *Coccidioides, Histoplasma, Blastomyces or Uncinocarpus spp*. (Fig. 17a). The lack of PL in the dimorphic fungi *Sporothrix, Histoplasma and Coccidioides* spp. has been previously shown by Teixeira, *et al*,[133] who suggested that the lack of PL could be suggestive of an adaptation away from the degradation of plants, towards an ecological niche within mammalian hosts. The lack of GH28 and GH93 family CAZymes in *Coccidioides spp*. would also support this hypothesis.

A family GH92 enzyme, alpha-1,2-mannosidase (CIMG_10116), was absent from the reference proteomes of both available *C. posadasii* strains, but was encoded for in all 4 available *C. immitis* strains and all other fungi evaluated. An orthologue search of this gene against sequenced *C. posadasii* genome annotations revealed a gene with 41% identity (E value of 1e-[15]) to a 164 amino acid protein in *C. posadasii* strain Silveira (CPSG_08711; conserved hypothetical protein), which lacks the GH92 functional domain. Alpha-1,2-mannosidase is a conserved protein, which in other fungi has a much longer length of >750 amino acids, so the absence of the GH92 functional domain in only *C. posadasii* could potentially be due to a genetic loss or due to incomplete genome sequencing/ annotation errors. However, if the GH92 domain is truly missing from *C.*

posadasii and not *C. immitis*, this would be a species-specific CAZyme difference between *C. immitis* and *C. posadasii*, whereas families AA12, GH28 and GH93 may indicate *Cocccidioides* genus-specific differences in the CAZomes.

Out of the 89 CAZyme families encoded in the 7 combined proteomes of *Coccidioides spp.* in the Uniprot database, 60 (72.3%) of the protein products were detected experimentally in this study. This indicates that numerous CAZymes are actively being translated *in vitro* and *in vivo* and suggest their importance throughout the growth conditions tested. Although only 32 (36.0%) of the total CAZyme families were identified in *Cocccidiodies spp.*-infected lung tissues, many CAZymes have signal peptides which allow for their secretion outside of the cell which could cause them to have been missed during tissue processing and laser capture. However, 63 (70.8%) of all CAZyme families encoded within the reference proteomes were represented by proteins detected from lysates of *in vitro* cultures.

In addition to comparing CAZymes that were experimentally detected against the *Coccidioides spp.* CAZymes in the reference proteomes, we evaluated relative CAZyme protein abundance differences in mycelial and spherule growth forms in multiple media and compared it the relative protein abundances *in vivo*. Intriguingly, a number of CAZyme protein abundance differences were observed between the spherule and mycelial form cultures. Endochitinase1, Beta-hexosaminidase1, Alpha-mannosidase, Mannosyl-oligosaccharide alpha-1,2-mannosidase, and Beta-1,3-glucanase enzymes were

significantly higher in the mycelial form, whereas Choline Oxidase, NADP-dependent

leukotriene B4 12-hydroxydehydrogenase (LTB$_4$D), and Beta-glucosidase 5 (BGL5)

were significantly higher in the spherule form (Fig. 18a).

These findings are of particular interest because Endochitinase1, also known as

complement-fixation "CF" antigen, is one of the 2 main Coccidioidal antigens currently

used in immunodiffusion diagnostics for Valley Fever. The mean abundance of CF

antigen was approximately 70-fold higher in the mycelial form than in spherule form

cultures or *Coccidioides spp.*-infected human lung tissues. Since spherules, not mycelia,

are the predominant fungal form that the immune system would most likely encounter, it

is curious that antibodies against a predominantly mycelial antigen are used in diagnosis

and monitoring of patients with coccidioidomycosis. It is possible (although unlikely)

that CF and TP are more immunogenic than other antigens predominantly expressed by

spherules. Armed with this knowledge, one might choose an antigen for serological

diagnostics that is predominantly expressed in spherules.   The other antigen used in VF

immunodiffusion is a 120kDa beta glucosidase (BGL2), also known as tube precipitin

"TP" antigen. Similarly to CF, the abundance of BGL2 was 8-fold higher in mycelial

cultures than in spherule cultures, however the difference was not significant by 2-way

ANOVA.

Of the CAZyme protein abundances significantly increased in the spherule form,

both Choline Oxidase and BGL5 were increased approximately 3-fold greater in the

spherule form than in the mycelial form. BGL5 is a 56kDa beta-glucosidase which shares enzymatic function but does not share sequence homology to the BGL2 antigen published as TP antigen.[134] Little is known about the specific role of beta-glucosidase or choline oxidase in fungal infections. LTB$_4$D abundance was also observed 4.5 fold higher in abundance in the spherule form and in *Coccidioides spp.*-infected LCM lung tissues than in mycelial form cultures. LTB$_4$D is involved in the inactivation of the chemotactic factor, leukotriene B4, an important early recruiter of neutrophils and effector CD8+ T cells.[135] In a murine study of *A. fumigatus* pulmonary infection, host-derived leukotriene B4 was determined to be crucial for host resistance to infection, as leukotriene B4 knockout mice had impaired fungal clearance.[136] It is therefore possible that fungal inactivation of host leukotriene B4 by LTB$_4$D could play a role in immunological dysregulation and promotion of fungal colonization.[137]

Also of interest are the media-specific CAZyme abundance profiles seen among the 6 different media formulations evaluated (Fig. 18a). Modified Converse with Tamol and RPMI with FBS are two of the most published growth media for *Coccidioides spp.* and numerous CAZyme abundance differences were seen between them. For example, GH55 (Beta-1,3-glucanase), which was significantly more abundant in the mycelial form, was 165-fold more abundant in RPMI with FBS than in Modified Converse with Tamol media. On the other hand, in spherule form cultures, RPMI with FBS produced roughly half the relative abundance of flavoprotein oxidoreductases of the AA6 family than

103

Converse with Tamol, and did not correspond to the abundance in LCM lung tissues. This highlights the importance of using *ex vivo* samples wherever possible and not relying solely on one growth medium when performing "Omics" studies.

Phase-dependent and media-specific *Coccidioides spp*. CAZyme protein abundance differences are intriguing and warrant further investigation. As CAZymes are responsible for the construction and degradation of glycans during and after protein translation, CAZyme presence or abundance could have a significant impact on the polysaccharide structures produced by the fungus.

In order to explore the possibility of *Coccidioides* genus-specific glycosylations, the N-glycomes of different fungi from infected human lung tissues were evaluated. Since glycans produced *in vivo* are most representative of the N-glycome produced within a host during infection, fungal-infected lung tissues were employed in this analysis. It is known that host factors like immunological defenses and a limited nutrient source can dramatically impact gene expression,[138,139] and we wanted to capture the N-glycome during this dynamic host-pathogen interplay. Glycans in culture lysates were evaluated for one of the most commonly used lab strains, *Coccidioides posadasii* strain Silveira, and were used for comparison purposes. The N-glycomes were evaluated for unique N-glycans, differences between monosaccharide and glycan motif abundances versus normal human lung tissue in each fungal group. The focus of this study was on *Coccidiodies spp*., however, the raw mass spectrometry files and analyzed lists used in

this study are freely available on the Pride Repository (Dataset Identifier: PXD012459) and Appendices, for other scientists who wish to delve deeper into the analysis of the data acquired for the other fungal species.

Although the glycan database used in this study contains over 10,000 published glycans, only ~33-43% of all unique m/z values in each fungal group were found in the database (Fig. 19). Unique MS1 m/z values were only evaluated in this study if there were MS2 spectra for that precursor m/z, giving support that the m/z were indeed glycans, not artifacts. Additionally, sample processing with PNGaseF and desalting though a PGC column selects for glycans. Therefore, the precursor m/z values that were not found in the glycan database are likely to be glycan structures that are not yet characterized and published in the literature. Figure 19 exemplifies the argument that there are more unknown glycan structures than known at this point in time. Future studies using fluorophore-labeled mass spectrometry or nuclear magnetic resonance imaging could help further characterize structures of these unknown/unpublished glycans.

Of the known glycan structures found in the database, monosaccharide abundances were calculated and glycan motifs were pulled from the GlyTouCan repository.[140] Monosaccharide compositions and glycan motifs can inform of glycan function and aid discovery of interacting proteins. Lectins, glycan-binding proteins and antibodies can bind to specific glycan motifs and residues. Knowing the prevalent motifs produced by an organism can be useful for immunoprecipitations or

105

immunofluorescence, which could aid in the diagnosis or further characterization of an organism. In this study, GlcNAc and galactose were significantly greater in abundance in coccidioidal spherules than in normal human lung tissue. This corresponds with our previously published immunohistochemistry data indicating that succinilated wheat germ agglutinin (sWGA) and *Griffonia simplicifolia* II (GSLII) lectins, which bind to GlcNAc dimers/trimers and do not bind sialic acids, specifically bound to *Coccidioides spp.* spherules and not lung tissues.[64] A number of other motif and monosaccharide abundance differences between the other fungi tested and uninfected normal lungs can be seen in Figures 20 and 21. These differences could be targets for confirmation of unique motifs by lectin immunohistochemistry or microarrays.

Glycans play key roles in cell communication, protein networks and immunity and it has even been suggested that certain monosaccharides elicit different immune responses than others.[141] In mammals, mature glycans are typically capped with sialic acid residues, particularly when linked to an $\alpha(2,6)$ galactose, as exposed terminal galactose and GlcNAc produced by some fungi are immunogenic in humans. Of the four possible *Coccidioides spp.*-specific glycans with available structures, three had terminal galactose, one had a terminal GlcNAc and three had a core α-1,6 *fucose* at the *reducing end* of GlcNAc. Core α-1,6 *fucosylation is known in various cancers, but has not been previously attributed to fungi, as they have not been observed to have genes encoding known fucosyltransferases*. Fucose is a common sugar in animals, but is scarce in fungi.

106

There are no proteins in *Coccidioides spp*. currently annotated to contain

fucosyltransferases and a genomic analysis did not find any obvious orthologues. Three

of the unique glycan structures identified in only *Coccidioides spp*.-infected lungs were

also found in *in vitro* spherule cultures, which precludes them from being human host

glycans. An alternative to core fucosylation is core glycosylation by another deoxy

hexose with the same molecular weight, like rhamnose or fuculose. Chromatography of

acid hydrolysates of mycelial form *C. immitis* has been previously shown to contain

rhanmnose.[47] In fact, UDP-L-rhamnose biosynthesis was another enriched metabolic

pathway up-regulated in spherules in the meta-analysis study by Kirkland and Fierer.[128]

However, rhamnosyltransferases are yet to be attributed to fungi, even though there is

increasing evidence of rhamnose-containing glycans in fungi.[142,143] Future experiments

using immunohistochemistry of *Coccidioides spp*. with rhamnose binding lectins and

fucose binding lectins like Aleuria aurantia lectin that binds alpha-1,6 fucoses bound to

GlcNAc, would help clarify whether *Coccidioides spp*. do indeed have glycans that

contain rhamnose. Confirmation of novel glycosylations or novel glycosyltransferases in

Coccidioides spp. should be explored.

Although we were not able to directly link a CAZyme produced only in

Coccidioides spp. to a glycan structure produced only in *Coccidioides spp*., we identified

unique glycans only found in *Coccidioides spp*., suggesting opportunities for specific

diagnostic and treatment targets.

CHAPTER V

EVALUATION OF GENUS-SPECIFIC GLYCANS FROM HUMAN AND DOG

URINE AS DIAGNOSTIC BIOMARKERS FOR COCCIDIOIDOMYCOSIS

Abstract

There is a clinical need for a highly accurate, rapid antigen-based diagnostic for Coccidioidomycosis. The first step to achieve this goal is to identify candidate antigens and evaluate them as biomarkers of infection. In this Chapter, a methodology for enrichment of free glycans from urine is provided and evaluated. Free glycans from 77 human urines, 63 dog urines, and mycelial and spherule form cultures of *Coccidioides posadasii* strain Silveira were enriched to identify *Coccidoides*-specific glycan antigens excreted in urine. Limit of detection experiments using a 2AB-Man9 glycan standard diluted in normal urine indicated a sub-picomolar range of detection using the described methodology. Significance ranking and manual curation of glycan m/z allow the identification of six *Cocccioides*-specific glycans from VF positive human urines and 3 from VF positive dog urines were required to reach 100% diagnostic sensitivity and specificity. Additionally, a biomarker database of the 103 most significant glycan m/z peaks identified from the human training dataset was able to correctly diagnose 23 blinded urine samples with 93.33% sensitivity and 100% specificity. Finally, evidence is provided that anti-sera from rabbits immunized with glycan-enriched cultures produce

108

measurable anti-*Coccidoides*-specific glycan antibodies which have potential for use in lateral flow or MALDI-TOF diagnostic assay platforms.

Introduction

In the previous chapter, genus-specific Coccidioidal N-glycosylations were identified from *Coccidioides*-infected lung tissues. However, it was unknown whether their glycan subunits were released into peripheral blood or urine. Blood serum and plasma are highly complex, containing hundreds of proteins at any given time. As urine is a less complex medium and urinary antigen tests for other fungal polysaccharides exist, urine was chosen as the specimen of choice for this analysis.

Antigen-based lateral flow assay (LFA) diagnostics from urinary glycans are currently used to diagnose a number of respiratory infectious diseases; from Legionellosis, Histoplasmosis, Aspergillosis, and Pneumococcal pneumonia.[144] Urinary glycan antigen assays for Coccidoidomycosis are also available on the market, however these assays measure galactomannan, which is only clinically useful for human cases of severe or disseminated infections. Even in severe and disseminated cases where antigenuria is expectedly high, sensitivity is poor (~71%), and cross-reactivity with other fungal species is common. As previous reports indicate that the Coccidioidal cell wall is ~50-60% carbohydrate[17,18], and data from Chapter IV indicate a diverse array of genus-

specific glycans are produced by the fungus, the identification of genus-specific glycans in urine could provide better diagnostic performance.

The aim of the experiments described in this chapter was to enrich glycans in urine from VF and non-VF humans and dogs and to employ a screening funnel for which m/z peak identifications for glycans in non-VF urine were subtracted from m/z peak identifications for glycans in VF urine. The remaining glycans were then matched to m/z peak identifications from spherule supernatants. The *Coccidioides*-specific glycan biomarker candidates identified in this manner can be conjugated to immunoreactive proteins and used for animal vaccinations to produce anti-glycan antibodies. These *Coccidioides*-specific glycan antibodies could then be used to capture circulating glycan antigens in VF patient blood or urine prior to LC/MS, or applied in other antigen-based diagnostic assays like ELISAs, LFAs or even MALDI-TOF platforms.

Materials and methods

Urine Sample Collection

Human urine specimens were acquired from excess clinical specimens collected for routine standard of care at Mayo Clinic Arizona, under a previously approved IRB protocol with waived consent. Dog urine samples collected from Arizona were acquired from excess veterinary specimens collected for routine standard of care by Dr. Richard Stolper, DVM at the Scottsdale Ranch Animal Hospital. Non-endemic dog urine samples

were collected at the University of Missouri College of Veterinary Medicine by Dr. Angela Royal, DVM.

Samples had generally been stored in the clinical lab at refrigerated temperatures for 1-7 days before being frozen at -20°C or lower until analysis. VF status of humans was determined based on clinical evaluation using the Mayo Clinic Arizona multi-factorial criteria for the diagnosis of Coccidioidomycosis.[35] This diagnostic rubric includes the evaluation of patient symptoms, radiography, serology, histology and culture results, as no single assay or measure can currently be used alone to diagnose VF in humans. Positive VF status of dogs was defined as a positive IgG or IgM IDCF titer with clinical symptoms.

Fungal culture production

Mycelial and spherule cultures *of C. posadasii* strain Silveira were grown for 7 days *in vitro* in the following culture media; 1) Converse with Tamol, 2) RPMI+10%FBS, 3) RPMI+Survanta. Additionally, *C. posadasii* strain Silveira mycelia and mycelia of *Uncinocarpus reesii* (ATCC 34534), were grown in liquid cultures of glucose yeast extract (GYE) broth. Culture supernatants were filtered through a 0.45μ vacuum filtration device and frozen at -80°C until passing sterility checks. Sterility was assessed by plating 1/10 of total volume onto GYE agar plates and assessing mycelial growth after 3 weeks at room temperature.

111

Urine and fungal glycan enrichment- PGC SPE

Using a vacuum manifold, Hypersep Hypercarb 50mg SPE cartridges (Thermo 60106-303) were equilibrated by passing 1ml of 1M NaOH through the cartridge bed, followed by 2ml MilliQ water, 1ml 30% acetic acid, 2ml MilliQ water, 1 ml 50% ACN with 0.1% TFA Elution Solvent, 1ml 5% ACN with 0.1% TFA Wash Solvent. Either 500μl of urine or fungal culture supernatants were loaded onto the cartridge bed and the effluent was slowly collected by hand pipetting without allowing air to enter the cartridge bed. Effluent was passed through the SPE cartridge twice more. Cartridges were washed with 1ml MilliQ water and 1ml Wash Solvent. Purified and desalted urine glycans were eluted by passing 1ml of Elution Solvent and collected by pipetting into a clean 1.5ml eppendorf tube. Samples were speed vacuumed to dryness at 45°C and stored at -20°C until MS analysis.

Urine deglycosylation and enrichment- sWGA lectin

A bead bed volume of 400μl succinilated wheat germ agglutinin (sWGA) lectin (Vector Labs) was put into 2ml disposable chromatography columns (Pierce) and washed with 10 bed volumes of PBS prior to the addition of 1ml of urine. Lectin columns were allowed to bind to the urine for 1hr at room temperature by end over end rotation. The columns were then washed with 10 bed volumes of PBS prior to eluting with 500μl of 4M urea pH 3. Elutions were neutralized and the urea diluted with 500μl of 1M

112

potassium phosphate buffer (pH 8) prior to PNGase F (NEB) deglycosylation ON at 37°C. Free and deglycosylated N-glycans were then desalted and cleaned with PGC SPE as above.

Limit of detection evaluation

Glyko® 2-AB oligomannose 9 (2-AB Man-9; Prozyme; monoisotopic mass= 2002.7133) was either directly injected into the mass spectrometer or diluted in $500\mu l$ of normal donor urine prior to PGC SPE processing as above. Multiple dilutions were performed and the limit of detection was evaluated after duplicate technical runs.

Mass Spectrometry of Glycans

PGC SPE urine glycan pellets processed above, were reconstituted in $10\mu l$ of 0.1%FA in H_2O prior to centrifugation at 18,000rpm and loading into MS autosampler vials. LC-MS/MS was performed by loading $1\mu l$ onto a Dionex UltiMate® 3000 RSLC liquid chromatography (LC) system (Thermo, San Jose, CA) using a PepMap RSLC C18 2um, 75um x 25cm EASY-Spray™ column (Thermo). Glycans were separated using a 0.3 uL/min LC gradient comprised of 2%-90% mobile phase B in 0-120 min. Mobile phase A and B were 0.1% FA in water and acetonitrile, respectively. Eluting peptides were directly injected into a Orbitrap Elite Velos mass spectrometer (Thermo-Fisher) and ionized using collision-induced dissociation (CID) in positive ion mode. A "top 15″ data-dependent MS/MS analysis was performed (acquisition of a full scan spectrum followed

113

by collision-induced dissociation mass spectra of the 15 most abundant ions in the survey scan).

Rabbit Immunizations

Two New Zealand White rabbits were housed and immunized at Josman laboratories (Josman, LLC), using standard immunization protocols. An initial immunization was carried out at Day 0 with 500μg of *C. posadasii* strain Silveira spherule lysate, followed by a boost immunization with 300μg of the same lysate at Day 21. On Day 35, a second boost immunization was given containing 300μg of *C. posadasii* strain Silveira spherule lysate mixed with 100μl of spherule supernatant which was 15x concentrated using an Amicon 10kDa ultrafiltration device (Millipore) and 100μl of PGC SPE- enriched spherule supernatant glycans. Rabbit sera used in this study were collected on Day 0 (D0) prior to first immunization and on Day 42 (D42) post-immunizations.

Glycan ELISAs

It is known that polysaccharides less than ~10kDa do not efficiently bind to the hydrophobic surfaces of plastics.[145] To perform anti-glycan ELISAs, it was necessary to selectively biotinylate the reducing ends of *Coccidioides sp*. polysaccharides by reductive amination and coupling to biotin-LC-hydrazide (Thermo), using a previously published protocol.[146] This material was then able to bind to neutravidin-coated ELISA plates

114

(Costar 3590, Corning). In brief, 4.6mg of biotin-LC-hydrazide was dissolved in 70μl of dimethyl sulfoxide (DMSO) after heating for 1 minute at 65°C. This mixture was vortexed and 30μl of glacial acetic acid was added and the mixture was combinded with 6.4mg of sodium cyanoborohydride. A volume of 8μl of this mixture was added to dried polysaccharides from 50mL of *C. posadasii* strain Silveira spherule supernatant grown for 10 days at 40°C shaking at 180RPM in RPMI supplemented with 0.1% Survanta after enrichment on a 500mg PGC SPE column and reconstitution in 50μl of HPLC grade water. The reaction was carried out in a thermal cycler set to 65°C for 2 hr in a fume hood. The reductively biotinylated *Coccidioides sp.* glycans were purified immediately by cleaning on a 50mg PGC SPE column and stored at 4°C until use.

ELISA plates were coated with 10μg/mL of NeutrAvidin (Thermo) at 4°C overnight. The next day, 10μl/mL of the reductively biotinylated PGC-enriched *Coccidioides* glycans was added to the wells for 1 hr. at room temperature. Control antigens used included 2μg/mL of the glycoprotein Quiescin Sulfhydryl Oxidase 1 (QSOX1; recombinant protein produced in house) and 7μg/mL of N-acetylglucosamine derived bovine serum albumin (GlycNAc-BSA; Vector Labs), which were NHS-LC-biotinylated (EZ-link NHS-LC-biotin; Thermo). Wells were then blocked with Carbo-Free blocking buffer (Vector Labs) for 1hr prior to loading 2-fold dilutions of rabbit sera into the wells of the plate. A goat-derived anti-rabbit IgG (H+L) secondary antibody conjugated to HRP (Jackson Laboratories) was used at a 1:5,000 dilution. Color was

developed with TMB substrate (BD Biosciences), stopped with 0.1M sulfuric acid, and

read at 450nm on a spectrophotometer (Molecular Devices).

Glycan Identification

Glycan MS/MS spectra were analyzed using SimGlycan v.5.60 (Premier Biosoft;

Palo Alto, CA). Search parameters were underivatized, free glycans in positive ion mode

with H+ and Na+ adducts. A 1 Da ± 0.5 Da error tolerance was allowed. All glycosidic

and cross-ring options were set to "yes". Only precursor mass to charge ratios (m/z)

which also had MS2 spectra were kept for analyses.

Biomarker identification and statistical processing

Following identification of Coccidioidal glycans, MS-Bidet, an R script algorithm

developed by ASU scientists Dr. Reed Cartwright and his graduate student, Adam Orr,

was used to compare glycan datasets and provide statistical inference using a one-tailed

Fisher's Exact test. GraphPad Prism v8.0 was used for ELISA analyses and Microsoft

Excel for generation of Pareto charts.
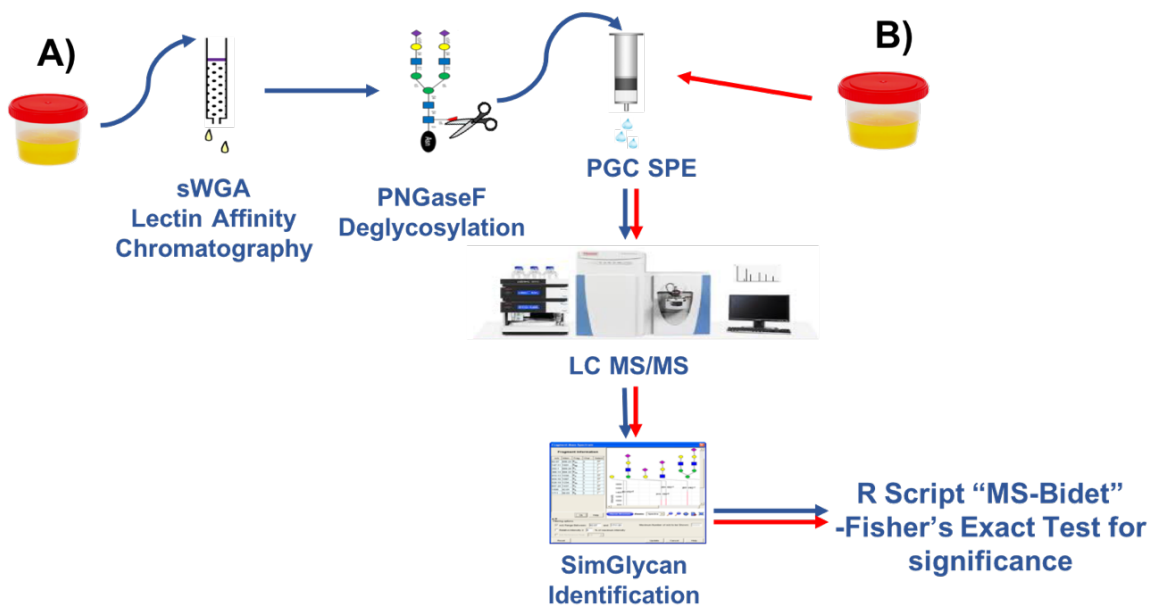
Figure 23. Urine glycan enrichment and identification methods used in this study.
Urine was initially processed using sWGA lectin affinity and deglycosylation of proteins
prior to porous graphitic carbon (PGC) solid phase extraction (SPE) of glycans (A; blue
line). However, processing was streamlined after evaluation studies to show more
abundant enrichment with PGC SPE alone prior to LC/MS and glycan identification with
SimGlycan software (B; red line).

Results

Human and dog samples used in biomarker training datasets

Urine samples used in this study for human and dog training datasets can be seen

in Table 3. There were a total of 77 human urines used in the human biomarker training

set for this study; 20 with confirmed VF, 49 endemic non-VF controls, and an additional

8 human urines from confirmed Histoplasma positive patients from Mayo Clinic

Rochester, which were used as non-endemic negative controls. Demographic information

117

for the Histoplasma positive patient urines was not available and is excluded from Table 3. There were 63 dog urines used in the dog biomarker training set for this study; 11 VF positive, 13 endemic non-VF controls and 39 non-endemic non-VF controls. Additionally, 3 technical replicates of 7 spherule form and mycelial form lysates from *in vitro* cultures of *C. posadasii* strain Silveira were used as *Coccidiodies* positive control fungal glycans and triplicate *Uncinocarpus reesii* culture lysates and supernatants were used negative control fungal glycans for both the human and dog biomarker training datasets.

		Total n (%)	Gender		Age		Clinical Course of Disease[a]			Serology (EIA, CF or ID)[b]		Antifungals?		
			Male	Female	<65	≥65	Acute	Chronic	Disseminated	Positive	Negative	Yes	No	Unknown
Valley Fever Positive	Human Urine	20 (29)	13 (65)	7 (35)	12 (60)	8 (40)	13 (65)	2 (10)	5 (25)	17 (85)	3 (15)	13 (65)	7 (35)	0 (0)
	Dog Urine	11 (17.5)	6 (54.5)	5 (45.5)	N/A	N/A	N/A	N/A	N/A	11 (100)	0 (0)	4 (36.4)	4 (36.4)	3 (27.2)
Valley Fever Negative	Human Urine	49 (71)	20 (40.8)	29 (59.2)	32 (65.3)	17 (34.7)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Endemic Dog Urine	13 (20.6)	10 (76.9)	3 (23.1)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Non-Endemic Dog Urine	39 (61.9)	22 (56.4)	17 (43.6)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 3: Human and dog urine donor demographics. Human urine samples from 20 confirmed Valley Fever patients and 49 endemic normal donors were used in this study. Dog urine samples from 11 Valley Fever positive dogs, 13 endemic negative dogs and 39 non-endemic negative dog urines were used.

119

Evaluation of sWGA versus PGC glycan enrichment

An initial evaluation on a small number of dog urine samples (n=9 VF positive, n=39 non-endemic VF negative dog urines) was performed using both lectin and PGC enrichment methodologies (Figure 23). This evaluation was carried out in order to determine the glycan enrichment method which would identify the greatest variety and number of glycans from urine, and inform of the method to use for the remaining urine samples. The 25 most frequently occuring m/z values in positive dog urines which were not in any negative dog urine, were assessed in Table 4.

		# of Observations by Enrichment Method (n=9)	
	Biomarker m/z	**sWGA lectin**	**PGC SPE alone**
1	425.863	1	7
2	496.736	5	4
3	484.521	4	6
4	650.759	5	3
5	718.335	2	5
6	826.647	1	6
7	546.577	0	7
8	570.673	2	5
9	565.669	3	5
10	543.744	5	2
11	747.782	2	5
12	611.008	2	5
13	927.387	2	5
14	620.775	3	4
15	1075.99	1	5
16	711.357	2	4
17	539.234	2	4
18	899.829	2	4
19	911.436	2	4
20	620.8	1	5
21	539.679	1	5
22	492.559	0	6
23	1020.203	0	6
24	485.739	3	4
25	575.281	2	4
	Total Observations	**53**	**120**

Table 4. Evaluation of the number of times each of 25 biomarker m/z values were observed in 9 VF-positive dog urines enriched with sWGA lectin or PGC SPE alone.

Three of the 25 (12%) m/z values were not present in any urines enriched by the

lectin method, whereas all 25 m/z values were present using the PGC enrichment method.

121

Of the 3 m/z values not seen using lectin enrichment, compositions for the 2 m/z values (546.577 and 1020.203) were not found in the glycan database and m/z value 492.559 was identified as KeggGlycan G04805 ($(Gal)_2$ $(GalNAc)_2$ $(GlcA)_2$ $(IdoA)_1$ $(SO3)_2$ $(Xyl)_1$), which does not contain the GlcNAc-GlcNAc binding motif required for sWGA binding.

The majority of m/z values were more prevalent when the PGC enrichment method was used (n=22, 88%). Finally, the 25 most prevalent m/z values were seen a total of only 53 times using the lectin method, whereas they were seen 120 times using the PGC enrichment method; an increase of 55.8%. Therefore, PGC enrichment alone was the method used for all other evaluations in this study.

Limit of detection

Picomolar amounts of 2-AB Man-9 was detected by C18 chromatography when directly injected into the mass spectrometer, as well as when diluted in normal donor urine and glycan-enriched through a PGC SPE column (Figure 24). The limit of detection of this method using 2-AB-Man9 was in the sub-picomolar range, as a concentration of 0.825pmol was seen with a mean relative intensity of 1.65×10^6. Below this concentration, signal to noise ratios were too poor to confidently identify this compound.
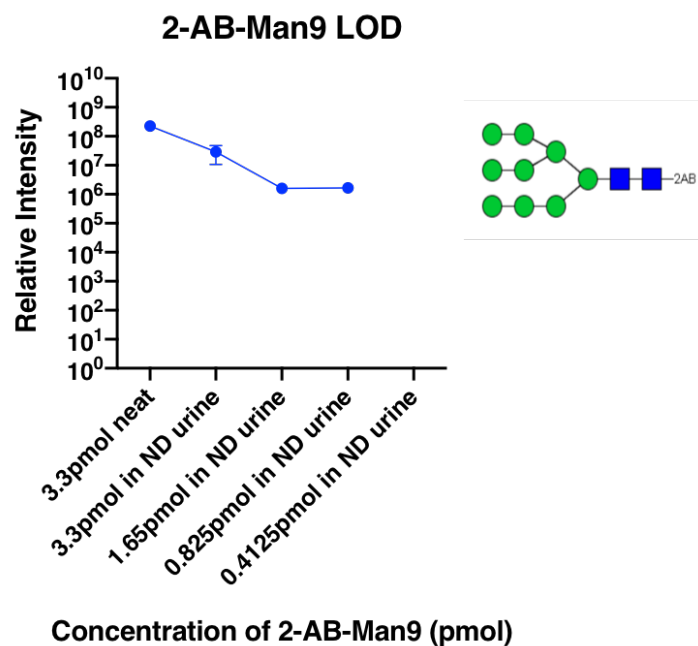
Figure 24: The limit of detection of 2-AB-Man9 diluted in normal donor (ND) urine and purified on a PGC SPE column is found to be 0.825pmol. Bars represent standard error of the mean (SEM) of duplicate experiments.

Human urine biomarker training dataset

There were 732,838 unique m/z values in the human urine dataset, of which 5,165 were significantly prevalent in the positive groups by Fisher's Exact Test (p<0.05). The m/z value with the greatest sensitivity was 726.274 (11/20, 55%), but it was present in 2 negatives (specificity of 96.6%). Without having a false positive (specificity of 100%), m/z 393.05 (p=4.35^{-6}) identified 8/20 (40%) positive patients. Combined, these 2 m/z provide 75% sensitivity and 96.6% specificity. Addition of m/z 803.272 raises sensitivity

to 95% and addition of m/z 908.386 makes sensitivity 100%. Although these 4

biomarkers combined provide the most accuracy with the fewest biomarkers, only m/z

803.272 was identified in Coccidiodes cultures (see Appendix E).

As the aim of this study is to find genus-specific diagnostic antigens, the data was

re-evaluated to only consider m/z values specific to VF patients (not in any control), and

which were simultaneously identified in Coccidioides cultures (specific to *Coccidioides*

sp. only). These parameters were set in order to ensure that the m/z values were

Coccidoides antigens, not host glycans. Using these assumptions, a human urine

biomarker training dataset containing 103 Coccidioides-specific m/z values was

generated. By manually curating this dataset, all 20 VF positive patients (100%

sensitivity) and all 57 VF negative urines were correctly categorized (100% specificity).

A pareto chart of the curated human biomarkers can be seen in Figure 25.
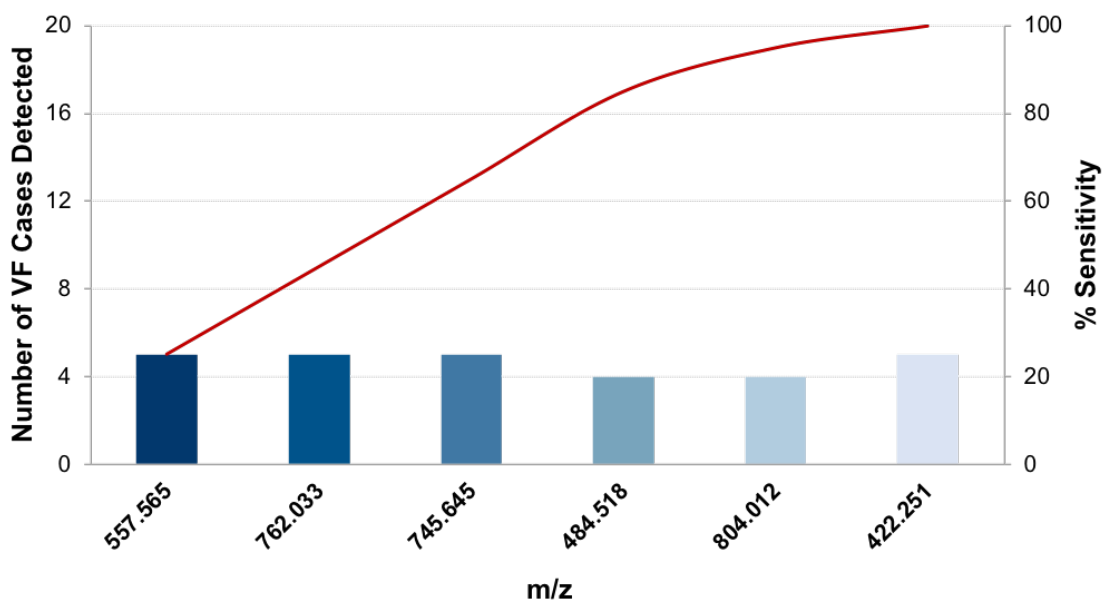
Figure 25: Pareto chart indicates that the combination of 6 biomarkers found in VF human urines is able to reach 100% sensitivity. Twenty VF positive human urines and 57 VF negative urines (49 normal donor urines and 8 *Histoplasma sp*. positive urines) were evaluated. Although each glycan biomarker was only detected in 4 or 5 of the 20 VF positive patients, sensitivity increased with increasing numbers of biomarkers added (red line). Sensitivity reached 100% with the addition of 6 biomarkers (m/z 557.565, 762.033, 745.645, 484.645, 484.518, 804.012, 422.251).

Human urine biomarker structures

Two of the 6 Coccidioides-specific urine biomarker m/z values did not confidently match the glycan database (m/z 762.003 and m/z 804.012). The remaining 4 m/z values matched published glycans, and their structures and MS2 fragmentation patterns can be seen in Figure 26.
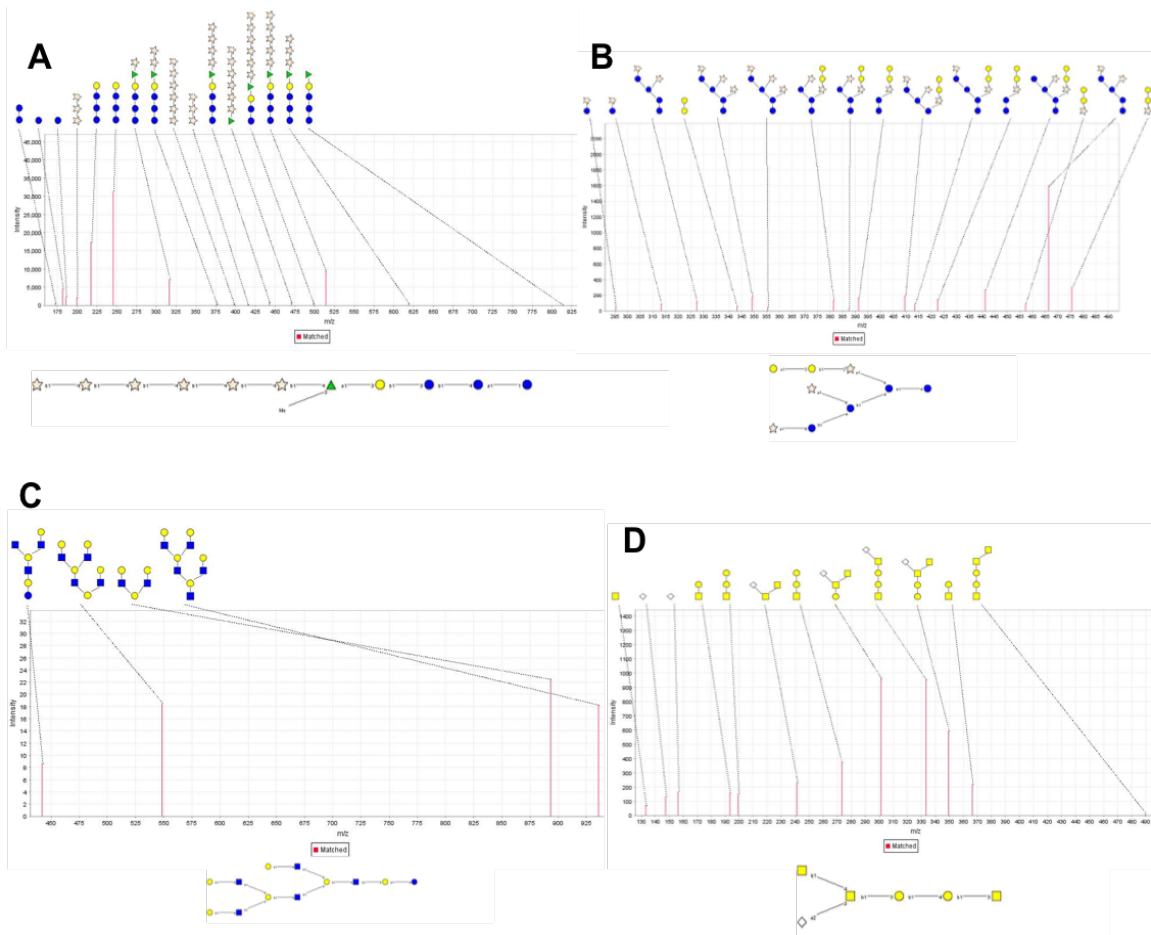
Figure 26: MS2 spectra of human biomarker glycans show confident matching with predicted glycan structures. Of the 6 *Coccidioides sp*. glycan biomarkers, 4 have known glycan compositions (A=m/z 557.565, B=745.645, C=484.518, D=422.251). Red lines indicate MS2 fragments which match the theoretical fragmentation pattern of the reference glycan. Complete parent structures for each glycan are found below each graph.

Dog urine biomarker training dataset

A total of 726,829 unique m/z values were identified in the dog urine dataset, of which 12,190 were significantly prevalent in the VF positive samples ($p<0.05$; see

126

Appendix F). The m/z value with the greatest sensitivity was 468.731 (10/11, 90.9% sensitivity), but it was present in 3 negative dog urines (specificity of 94.2%). This dog urine biomarker was also present in *Coccioides sp*. cultures.

Manual curation of the data to consider only m/z values specific to VF dogs (not in any control), and which were simultaneously identified in Coccidioides cultures (specific to *Coccidioides sp*. only), three m/z values from the dog urine biomarker training datset (m/z 554.246, 501.25 and 935.41) provide 100% sensitivity and 100% specificity in the diagnosis of all 11 VF positive dogs (Figure 27).
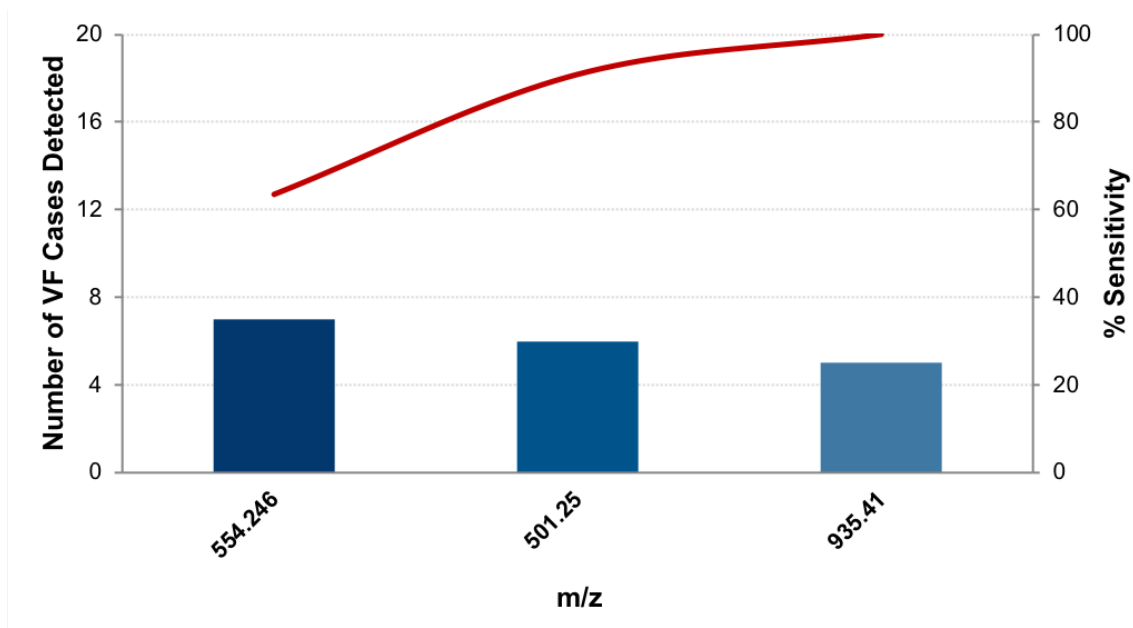
Figure 27. Pareto chart indicates that the combination of 3 biomarkers found in VF dog urines is able to reach 100% sensitivity. Eleven VF positive dog urines and 52 VF negative urines (39 non-endemic normal urines and 13 endemic negative urines) were evaluated. Although each glycan biomarker was only detected in between 5-7 of the 11 VF positive dogs, sensitivity increased with increasing numbers of biomarkers added (red line). Sensitivity reached 100% with the addition of 3 biomarkers (m/z 554.246, 501.25 and 935.41).

Dog urine biomarker structures

All 3 of the Coccidioides-specific dog urine biomarkers had confident matching

glycan structures in the database. Their structures and MS2 fragmentation patterns can be

seen in Figure 28.

128
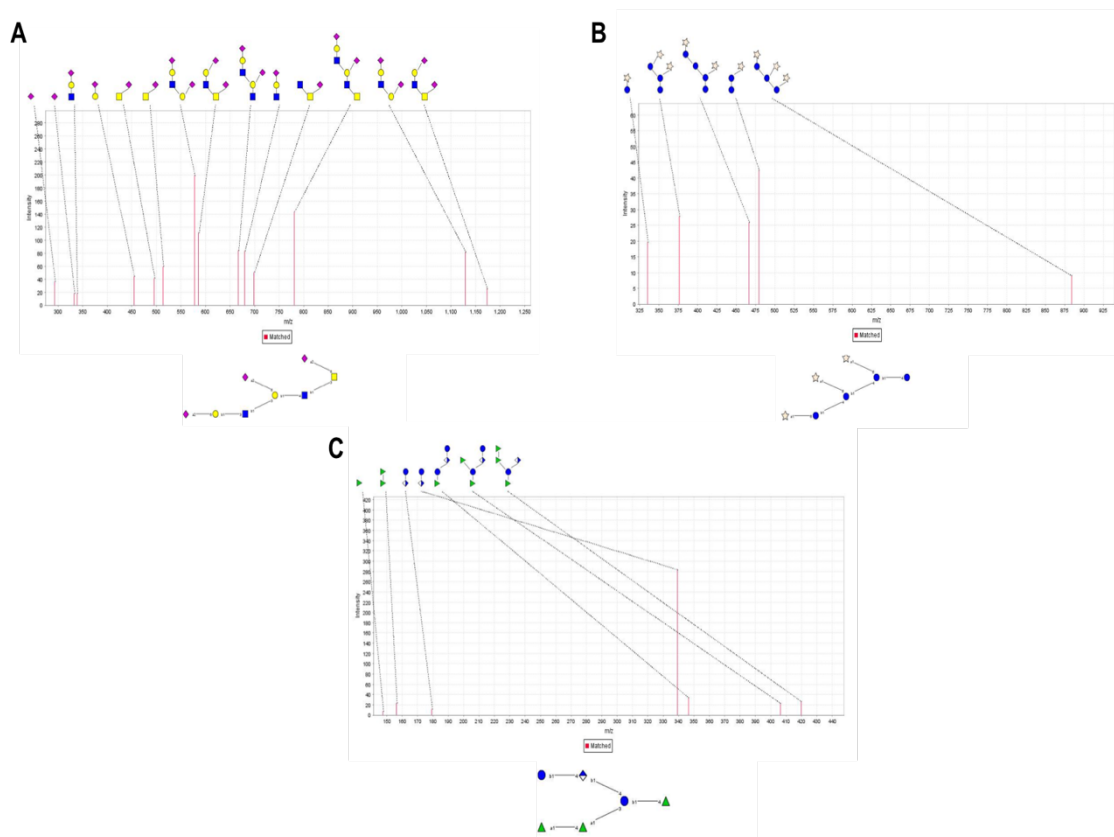
Figure 28: MS2 spectra of human biomarker glycans show confident matching with predicted glycan structures from dog urine. Of the 3 *Coccidioides sp.* glycan biomarkers in dog urines, all 3 have known glycan compositions (A=m/z 554.246, B=501.25, C=935.41). Red lines indicate MS2 fragments which match the theoretical fragmentation pattern of the reference glycan. Complete parent structures for each glycan are found below each graph.

Blinded human urines prospective and retrospective samples

Twenty three human urine samples were blinded by Dr. Thomas Grys of Mayo Clinic, by providing alphabetically lettered urine samples for evaluation. These urines contained 9 urines which were used in the human urine training dataset (39.1%) and 14

urines which had not previously been tested (60.9%). The human urine biomarker dataset

containing 103 m/z values was used to categorize patients from the blinded urines. If any

of the 103 m/z values were identified in a blind urine, it was categorized as a VF positive

patient. The biomarker dataset was able to correctly diagnose patients from all clinical

courses with high accuracy. There was one false positive noted and one indeterminant

urine, providing a sensitivity of 93.33% and specificity of 100%. Results from this

blinded study can be seen in Table 5.

A

	FALSE NEG	TRUE NEG
TRUE POS 14	**FALSE POS** 0	
1	8	

B

Correct?	Blinded Letter	Result (VF Y/N)	True VF (Y/N/IND)?	Original ID	Serology	Clinical Course
✓	A	Y	Y	U89	+	DISSEMINATED
✓	B	N	N	U187	NA	NA
✓	C	Y	Y	U109	+	ACUTE
✓	D	N	N	U175	-	RESOLVED
✓	E	Y	Y	U186	+	ACUTE
✓	F	N	N	U129	NA	NA
✓	G	Y	Y	U184	+	ACUTE
✓	H	N	N	U114	NA	NA
✓	I	N	N	U128	NA	NA
✓	J	Y	Y	U188	+	CHRONIC
✓	K	Y	Y	U178	+	CHRONIC
✓	L	Y	Y	U176	+	RESOLVING
✓	M	Y	Y	U181	+	RESOLVING
✓	N	N	N	U106	+	RESOLVED
✓	O	Y	Y	U185	+	RESOLVING
X	P	N	Y	U179	-	CHRONIC
✓	Q	Y	Y	U1	+	DISSEMINATED
✓	R	N	N	U183	+	RESOLVED
✓	S	Y	Y	U80	+	ACUTE
?	T	Y	IND	U177	+	IND
✓	U	N	N	U182	NA	NA
✓	V	Y	Y	U110	+	ACUTE
✓	W	Y	Y	U180	+	ACUTE

C

Statistic	Formula	Value	95% CI
Sensitivity	$\dfrac{a}{a+b}$	93.33%	68.05% to 99.83%
Specificity	$\dfrac{d}{c+d}$	100.00 %	63.06% to 100.00%
Positive Likelihood Ratio	$\dfrac{Sensitivity}{1-Specificity}$		
Negative Likelihood Ratio	$\dfrac{1-Sensitivity}{Specificity}$	0.07	0.01 to 0.44
Disease prevalence	$\dfrac{a+b}{a+b+c+d}$	65.22% (*)	42.73% to 83.62%
Positive Predictive Value	$\dfrac{a}{a+c}$	100.00% (*)	
Negative Predictive Value	$\dfrac{d}{b+d}$	88.89 % (*)	54.63% to 98.15%
Accuracy	$\dfrac{a+d}{a+b+c+d}$	95.65% (*)	78.05% to 99.89%

Table 5: Blinded human urine trial using prospectively collected samples. Human urine samples from 23 individuals (A) were tested against the database of the top 103 most significant biomarker candidates generated from the training dataset. Any matching m/z value found in a blinded sample was deemed a positive result (B). Using these criteria, a sensitivity of 93.3% and specificity of 100% was achieved (C).

131

Anti-Coccidioides glycan ELISAs

Two New Zealand White rabbits were immunized with PGC-enriched *Coccidoides sp.*

glycans and evaluated for anti-*Coccidoides sp.* glycan antibodies ten days post-

immunization by ELISA (Figure 29). Rabbits produced anti-*Coccidoides sp.* glycan

antibodies which are minimally cross-reactive with a human tumor-derived glycoprotein

and GlcNAc-BSA (ie. chitin-BSA).
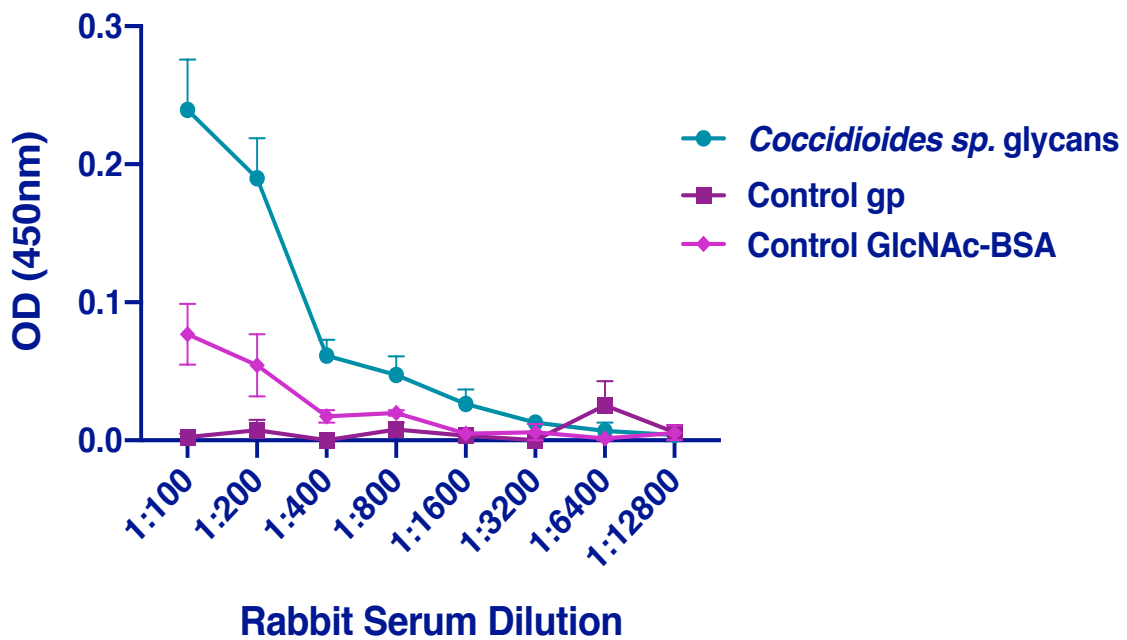
Figure 29: Rabbits immunized with enriched *Coccidioides sp.* polysaccharides produce antibodies specific to *Coccidioides sp.* glycans after 42 days. Following subtraction of pre-immune values, ELISA titration curves of serum from 2 immunized rabbits indicate antibodies are formed against enriched *Coccidioides sp.* glycans after 42 days, and these rabbit antibodies are minimally cross reactive with a control glycoprotein and GlcNAc-BSA. Error bars represent standard error of the means (SEM) of 2 rabbits.

132

Discussion

This study presented in this chapter provides evidence that *Coccidoides*-specific glycan antigens are released into both human and dog urine, and these glycans can be used as diagnostic biomarkers of active infection. These glycan antigens were present in patients regardless of clinical course or use of antifungals.

An initial evaluation of lectin versus PGC SPE enrichment clearly indicated that the PGC SPE enrichment identifies the greatest number and variety of glycans from urine. This observation is likely due to the fact that lectins are motif-specific, and in the case of sWGA, is specific to β-d-GlcNAc-(1,4)-β-d-GlcNAc. On the other hand, PGC SPE is not motif-specific, enriching by charge and polarity of the analyte. Lectins are thus more specific, but PGC SPE enrichment alone was more sensitive, faster and less expensive as an enrichment procedure for initial identification of total urine glycans. PGC SPE was therefore the preferred choice for glycan enrichment.

An unresolved question regarding the use of PGC SPE extraction of urinary glycans is what minimum concentration of urinary glycans are needed to be detected by the outlined mass spectrometry workflow after enrichment. Limit of detection experiments were therefore performed using a 2AB-labelled glycobiology standard, which was diluted in normal urine until extinction of mass spectrometry signal. This methodology was capable of detecting 2AB-Man9 as low as 0.825pmol (1.65 ng) in 0.5ml of normal urine. Sub-picomolar detection of glycans in urine is highly sensitive

133

and comparable to limit of detection studies without enrichment procedures.[147] However, different glycans will ionize or be PGC-enriched different than 2-AB Man9, depending on charge and size. Thus not all urinary glycans will be detectible at the same concentration. Regardless, an approximation that the LOD is in the picomolar/ sub-picomolar range is useful information which enhances our understanding of the capabilities of this methodology. As a comparison, Aspergillus EIAs detecting 1ng/ml of urine galactomannan are diagnostic.[148]

A total of 77 human urines and 63 dog urines were used in the training datasets to identify candidate glycan biomarkers which are *Coccidioides*-specific. M/Z values that were identified in *Coccidioides sp*. cultures and not in any negative urines or non-*Coccidiodes* fungal cultures were considered possible VF glycan biomarkers. These parameters were established to ensure that the m/z values were *Coccidioides*-specific antigens and not host glycans. Although these parameters could have removed potential biomarkers with greater sensitivity than the ones chosen, the purpose of this study was to determine *Coccidioides* genus-specific glycans in urine for use as an antigen-based diagnostic; we wanted to avoid false positives. *Coccidioides*-specific glycan biomarkers identified in human urines were ranked by statistical significance using Fisher's Exact Test; a way to test the association between two categorical variables (*ie*. VF or non-VF). The ranking of binary data (m/z present/ not present) using this test allowed for the identification of glycan m/z which are more frequently present in VF patients, not

controls. The outcome of applying this data processing method was a list of the "top 103 biomarkers", which was a list of all biomarkers with significance $p<0.001$.

The "top 103" biomarkers derived from the human training dataset were then used to evaluate a blinded mix of 14 prospectively collected, previously untested urines and 9 previously tested human urines. The presence of any m/z from the "top 103" list in any of the blinded urines triggered the categorization of that samples as a VF positive urine. After blind evaluation, an excellent sensitivity (93.33%) and specificity (100%) was achieved, highlighting the utility of this methodology in the diagnosis of VF. Due to limitations on the number of VF positive dog urines available, a blinded set containing previously untested dog urines was not possible. Future efforts to enroll additional veterinary clinics will be useful.

A minimum number of 6 *Coccidioides*-specific glycans from human urines and 3 *Coccidioides*-specific glycans from dog urines were needed to identify 100% of VF-positive urines. This manually curated set of glycans do not represent the only possible combinations to get to 100% sensitivity and specificity, but were chosen by ranked significance. Fewer biomarkers were required to identify 100% of VF-positive dogs (n=3) than VF-positive humans (n=6) in the respective training datasets. This may be a factor of the lower number of VF positive dog urines available (n=11) than VF positive human urines (n=20). Use of the Random Forest machine learning algorithm could improve biomarker classification and identification of patients in future.

135

The 6 curated *Coccidiodes*-specific glycan biomarkers from humans and 3 curated *Coccidiodes*-specific glycan biomarkers from dogs show diversity in structure and monosaccharide compositions (Fig. 26 & Fig. 28). Only 1 glycan structure from the curated human urine biomarkers (Fig. 26C) and 1 glycan structure from the curated dog urine biomarkers (Fig 28A) contained GlcNAc, but none contained the motif β-d-GlcNAc-(1,4)-β-d-GlcNAc. This observation supports the use of a broad glycan enrichment procedure with PGC SPE rather than a motif-specific lectin, as a diverse range of Coccidoidal glycans are shed into the urine. None of these curated glycan structures look like classical N-glycans, which is expected, as these are free glycans, not glycans cleaved from proteins using PNGaseF. However, as these are not enzymatically cleaved N-glycans, the statistically significant glycans present in urine in this study do not match the glycan structures identified from lung tissues in Chapter IV. Further experiments to isolate all urinary proteins for deglycosylation with PNGaseF could provide this comparative data.

Interestingly, one of the 6 curated human biomarkers was also present in the 3 curated dog urine biomarkers (Fig. 26B & Fig. 28B). Dog urine glycan Figure 28B ($(Glu)_4 (Xyl)_3$) is part of the same core glycan structure as human urine glycan in Figure 26B ($(Glu)_4 (Xyl)_3 (Gal)_2$), but without the antennary branch galactoses. It is possible that monoclonal antibodies produced to the dog glycan will also bind to the human glycan structure, allowing the generation of a simultaneous dog and human biomarker detection

136

antibody.

The lack of greater overlap between human and dog urinary glycans is not surprising, as previous evaluations of urinary galactomannan testing which provides ~71% sensitivity in humans, was only 3.5% sensitive in a panel of 60 dog urines.[26] It is likely that a different set of Coccidioidal glycans are released into the urine of dogs than humans, as host enzymes and urinary tract microflora differ between the host species,[149] generating different metabolite profiles.[150]

It is possible that the use of non-derivitized glycans in their free state precluded the visualization of additional glycan biomarkers. Permethylated glycans are known to offer better ionization and better chromatographic separation on C18 columns.[151] Permethylation involves the addition of a methyl group to all hydroxyl and N-acetyl groups, increasing the hydrophobicity of the molecule. However, this modification is labor intensive and therefore not conducive to a high throughput workflow. Additionally, it is not compatible with all classes of glycans. Identification of non-permethylated glycans in their native form were therefore pursued in this study. Alternative methods frequently employed to increase ionization efficiency of glycans include the addition of salts that provide sodium and ammonium ion adducts.[152] As urine naturally contains these ions, salt adducts are innately added to the glycans within urine. Future work aiming to identify urinary glycans with MALDI-TOF might be performed with permethylated glycans, as permethylated glycans are more stable during MALDI ionization.

Alternatively MALDI-TOF or LFAs could be used to identify glycans after antibody enrichment with anti-Coccidoidal glycan monoclonal antibodies formed against the glycan biomarkers identified herein. Rabbits immunized against PGC-enriched Coccidioidal glycans were shown to produce antibodies to PGC-purified glycans from *Coccidoides* culture supernatants. There was minimal cross reactivity of the rabbit sera with human glycoprotein QSOX1, showing specificity to fungal glycans. There was, however, a small amount of cross-reactivity to GlcNAc-BSA, which could be due to rabbit antibodies being formed to GlcNAc- containing glycans. However, as the majority of binding is not attributable to this chitin motif, the rabbits are primarily producing antibodies to other glycans and motifs. It is clear that the rabbits are producing some *anti-Coccidioides sp.* glycan antibodies, and we should expect higher antibody titers after subsequent antigen boosts. However, it remains to be seen if the antibodies will be specific to *Coccidioides* species glycans. A number of antibody screening steps will be needed to ensure cross-reactivity to human urine glycans or other fungi does not occur. HPLC fractionation of individual glycans will be required prior to screening of monoclonal antibodies. Regardless, this is proof of principle that anti-Coccidioidal glycan antibodies can be formed in rabbits. Further boosts with Coccidioidal glycans reductively conjugated to CRM197, an FDA-approved non-toxic 58kDa Diptheria protein, should enhance this response and allow for the production of monoclonal antibodies to these glycans. Coccidioidal glycan biomarker capture with monoclonal antibodies are also

likely to improve sensitivity and enrichment above that achieved by PGC SPE, allowing

the potential for greater diagnostic accuracy.

CHAPTER VI

DISCUSSIONS

Valley Fever is estimated to cause ~15-30% of all community-acquired pneumonias in the highly endemic Greater Phoenix and Tucson areas in Arizona.[153,154] Throughout the United States in 2017, there were 14,364 cases reported to the CDC, however low testing rates indicate that cases are underreported.[155] The lack of quick and accurate diagnostic assays for coccidioidomycosis during the acute phase of infection is one possible reason for low reporting rates. Current diagnostic platforms for acutely ill patients measure patient antibody responses to infection, which often take several weeks to develop, rather than measuring fungal antigens circulating in the blood or urine. An accurate antigen-based diagnostic is a much needed assay for coccidioidomycosis. Such a test would allow clinicians to determine current from previous infections, allow for faster and more accurate treatment and has the potential to increase reporting, as well as reduce patient morbidity and associated healthcare costs.

Protein biomarkers are the antigen type most commonly used in diagnostic platforms. As such, proteins would be attractive targets for antigen-based diagnostics for Valley Fever. However, individual protein biomarkers for antigen-based diagnostics for VF have not yet successfully been identified and published in the literature. Preliminary strategies employing sWGA and GSLII lectins or anti-coccidioidin polyclonal antibodies

to enrich coccidioidal proteins prior to mass spectrometry were not able to reliably identify *Coccidioides*-specific proteins or peptides in acutely ill patients. One possibility is that the fungal proteins are enzymatically cleaved by endogenous proteases and are quickly removed from circulation. In fact, no publications regarding protein or peptide biomarkers for any fungal mycosis were found in a literature search.

In Chapter II, the possibility that the low coccidioidal peptide identifications observed in the preliminary lectin and antibody enrichment experiments were due to low genome sequencing quality was explored. The *Coccidoides posadasii* strain Silveira genome was evaluated and re-annotated by using mass spectrometry-based proteomic evidence of proteins produced by the strain. The poor genome sequencing quality of *Coccidiodes immitis* and *posadasii*, combined with proteomic re-annotation of their genomes,[125] led to the conclusion that the proteomes of the two species are more similar than previously estimated. Around 12% of the strain Silveira genes identified were missing from the annotated genome, but were annotated in at least one of the other 6 published *Coccidioides spp*. genomes. The in-house creation and use of a proteomic database which includes all sequenced reference proteomes and newly annotated peptides allowed for improved protein identification from biological samples. However, use of the new database did not increase protein biomarker discovery when the preliminary raw dataset was re-analyzed. Therefore, lack of identification of reliable plasma and/or urine

141

peptide/protein antigens using lectin and antibody enrichments were thereby not due to annotation errors.

An alternative strategy for the identification and isolation of coccidioidal protein antigen biomarkers was therefore needed. As lung tissues are highly vascularized with blood and lymphatic capillaries, the most abundant proteins produced by coccidioidal spherules during growth and rupture, are likely to drain into the blood and possibly urine. In Chapter III, the most abundantly present proteins in *Coccidioides spp.*-infected human, dog and mouse lung tissues were thus explored. LCM of spherules followed by mass spectrometry allowed the identification of 807 coccidioidal proteins. From the top 100 most abundant proteins, 28 were identified as *Coccidioides spp.* biomarker candidates, as they do not share sequence orthology with any human protein. Twenty five of these 28 biomarkers are conserved in fungi. The remaining 3 biomarker candidates are also potential *Coccidoides*-specific biomarkers, as they also do not share sequence homology to any other pathogenic fungus or microbe. These 3 Uncharacterized proteins are small (<45kDa) secreted proteins that are genus-specific, making them the most attractive biomarker targets of the 28. As they are uncharacterized proteins, future experiments to characterize their function could be performed. Additionally, the generation of antibodies to these proteins could be used in future immunoprecipitation experiments from infected patient samples to selectively enrich for these proteins in blood or urine. Alternatively, they could be evaluated as vaccine candidates in mouse models.

The 28 candidate biomarker proteins identified from infected lung tissues are all potential targets for antigen-based VF diagnostics, however, only 2 of the 28 proteins (Peroxisomal Membrane Protein and Uncharacterized protein CIMG_09001) have been identified in biological specimens in the literature.[112] Since these proteins were only identified in the BALF of artificially-infected animals, it is still unknown whether these protein antigens could be isolated from naturally-infected human patients. A central question that arose from our studies is why coccidioidal proteins are so difficult to isolate from blood or urine. Different enrichment technologies, like microfluidic chips or nanotraps, or lectins and antibodies different from those tested, may be required to isolate low abundance transiently circulating peptides.

While searching for protein biomarkers from LCM lung tissues, a high abundance of CAZymes was noted in the proteomic data. It is known that CAZymes, like Endochitinase 1 (CTS1), are used in serology-based diagnostics and are important for virulence.[156] However, little is known about the glycans produced and acted upon by *Coccidioides spp*. CAZymes. By performing LCM capture and LC/MS glycomics and proteomics of infected lung tissues, the coccidioidal CAZome and glycome were experimentally determined. This is the first study to provide experimental evidence of CAZymes, their abundances and the identification of the spherule glycome from *in vivo* samples. The significantly greater abundance of Choline oxidase, Beta-glucosidase 5 and LTB$_4$D in the spherule phase indicates these CAZymes are not only potential biomarker

targets, they could also be targets for gene knockouts to assess virulence, or be used as therapeutic targets using small molecule inhibitors.

The identification of unique glycans found only in *Coccidioides spp.*-infected humans and in *in vitro* lysates, but not patients infected with other common fungal infections, suggest possible genus-specific glycosylations. Genus-specific glycosylations are corroborated by the literature in other fungi, including *Aspergillus spp.*,[67,118] yet have not been previously identified in *Coccidioides spp.* Interestingly, three of the four significantly present *Coccidioides*-unique glycan structures had core fucosylation motifs. These genus-specific motifs provide opportunities for specific diagnostic and treatment targets. The application of lectins specific to these motifs could provide confirmatory evidence of the specificity of these motifs in only *Coccidioides spp.*, and also have the potential to be used for patient sample enrichment of coccidioidal antigens. Although specific glycan structures found solely in *Coccidioides spp.* were identified herein, they could not be directly linked to a particular CAZyme or set of CAZymes. Generation of targeted CAZyme deletion mutants followed by glycomics analysis could provide this data.

The use of polysaccharides as antigens for fungal diagnostics is not novel. They have been used for years to diagnose a multitude of fungal mycoses, including *Histoplasma spp.*, *Aspergillus spp.*, *Candida spp.* and even *Coccidioides spp.* A number of companies provide antigen-based assays for analysis from a range of biological

144

samples including BALF, CSF, blood and urine. However, the available assays measure BG or galactomannan, which is not fungal genus-specific and not adequately sensitive enough to accurately diagnose acutely ill patients.[157] By identifying *Coccidoides* genus-specific glycosylations, a proof of concept for creating an antigen-based diagnostic specific for VF could be established.

The next goal was to determine if Coccidioides-specific glycans are present in biological samples from VF patients. Urine was chosen as the specimen of choice in these studies, as it is less complex than serum/plasma and is known to contain other fungal polysaccharides like BG and galactomannan. Urine collected from human and dog patients were evaluated for *Coccidioides*-specific glycans. The presence of glycans identified in multiple samples from human patients with different clinical courses, including acutely ill and seronegative patients matched glycans found in *Coccidoides sp.* cultures, but were not identified in any normal donor urine, *Uncinocarpus sp.* cultures or *Histoplasma* sp.- infected patient urines, suggests this approach is feasible. Adding credence to this is the ability to correctly classify positive from negative human VF patients in a blinded evaluation with 93.3% sensitivity and 100% specificity. Although the m/z values of the 3 most sensitive glycan biomarker candidates for dog urines were different than the m/z values of the 6 most sensitive human urine biomarkers, there was one core glycan structure shared between VF positive dogs and humans. This strongly

suggests that the fungus produces and sheds the same glycans into the urine of both humans and dogs.

The next logical step for this project was to immunize rabbits with PGC-purified coccidioidal culture glycans and evaluate immunological response to vaccination. The observation that rabbits can produce anti-*Coccidioides sp*. glycan antibodies after an initial priming immunization and only one boost is promising preliminary data. Carbohydrates are generally classified as T-cell independent antigens, as they activate B cells without MHC presentation to T cells.[158] This "V-region restriction" leads to the generation of antibodies which are of low affinity due to lack of T cell help. However, by conjugating glycans to peptides or full-length proteins, the glycoconjugate proteins recruit T-cell help, allowing the formation of antibodies to specific glycan epitopes.[159,160] A non-virulent mutant form of the diptheria toxin protein, CRM197, has been successfully used as a carrier protein for a number of FDA-approved polysaccharide vaccines, including Pfizer's Prevnar, the first Pneumococcal vaccine. By selectively conjugating the protein to only the reducing end of the polysaccharides using reductive amination with sodium cyanoborohydride, minimal glycan epitopes are lost.[161] The selective conjugation of CRM197 to only the reducing end of PGC-purified *Coccidioides sp*. glycans, could provide *Coccidioides*-specific IgG from immunized rabbits, for monoclonal antibody production. These antibodies, once screened for specificity on purified glycans, could be used to create various antigen-based diagnostics, including

146

lateral flow assays and ELISAs or used in immunoprecipitations prior to MALDI-TOF analysis.

Although a Valley Fever-specific diagnostic test was the main focus of this work, the technology, methods and data provided in this study could also be applied to develop a pan-fungal diagnostic assay. For example, if one were to re-analyze the data provided in Chapter IV, and only search for glycan structures which were identified in all fungal species but not in human lung tissues, those glycan structures could be explored for their potential to serve as universal biomarkers of fungal pneumonia. A non-invasive pan-fungal diagnostic assay would improve the current diagnostic and therapeutic landscape for invasive fungal disease (IFD). The most common fungal organisms associated with IFD are *Candida* and *Aspergillus spp.,* however infections with *Cryptococcus*, *Trichosporon*, *Malassezia*, *Fusarium*, *Scedosporium*, *Pseudallesheria*, *Rhizopus* and *Mucor spp*. are becoming more prevalent in North America and Europe.[162] Accurate and timely diagnosis of patients with IFD is required for successful treatment and clinical outcome, especially in cases of immunosuppression or malignancy.[163] As IFD can masquerade as a bacterial or viral infection, a diagnostic test which could rapidly "rule in" a fungal etiology would allow clinicians at primary care centers and emergency rooms to prescribe antifungal therapeutics at the first visit. However, due to the presence of a multitude of fungi which reside in the external environment and form part of our epidermal and internal flora, some quantity of fungal

147

antigens will always be in circulation. Therefore, it is likely that a pan-fungal diagnostic assay will require a quantitative platform, such as an EIA or Real-Time PCR. Indeed, this is currently what is required for Fungitell® BG assay. However BG is not produced by several fungal species, including *Cryptococcus*, *Rhizopus* and *Mucor spp*., thus a pan-fungal diagnostic is still required.

The results of this research provides evidence that *Coccidioides* genus-specific glycosylations have potential as antigens in Valley Fever diagnostic assays. Production of monoclonal antibodies to these glycan biomarkers is the crucial next step to developing a point of care anti-glycan antigen-based diagnostic for Valley Fever. With the limited sensitivity and specificity of the current serological and antigen-based diagnostics for coccidioidomycosis, the complementation of these tests with a genus-specific anti-glycan point of care test has the potential to allow clinicians to treat patients more rapidly, reduce diagnostic and healthcare costs, and even save lives.

REFERENCES

1   Hector, R. F. & Laniado-Laborin, R. Coccidioidomycosis--a fungal disease of the Americas. *PLoS Med* **2**, e2, doi:10.1371/journal.pmed.0020002 (2005).

2   Centers for Disease, C. & Prevention. Increase in reported coccidioidomycosis-- United States, 1998-2011. *MMWR Morb Mortal Wkly Rep* **62**, 217-221 (2013).

3   Lewis, E. R., Bowers, J. R. & Barker, B. M. Dust devil: the life and times of the fungus that causes valley Fever. *PLoS Pathog* **11**, e1004762, doi:10.1371/journal.ppat.1004762 (2015).

4   Shubitz, L. E., Butkiewicz, C. D., Dial, S. M. & Lindan, C. P. Incidence of coccidioides infection among dogs residing in a region in which the organism is endemic. *J Am Vet Med Assoc* **226**, 1846-1850 (2005).

5   Blair, J. E., Mendoza, N., Force, S., Chang, Y. H. & Grys, T. E. Clinical specificity of the enzyme immunoassay test for coccidioidomycosis varies according to the reason for its performance. *Clin Vaccine Immunol* **20**, 95-98, doi:10.1128/CVI.00531-12 (2013).

6   DiCaudo, D. J. Coccidioidomycosis: a review and update. *J Am Acad Dermatol* **55**, 929-942; quiz 943-925, doi:10.1016/j.jaad.2006.04.039 (2006).

7   Bried, J. M. & Galgiani, J. N. Coccidioides immitis infections in bones and joints. *Clin Orthop Relat Res*, 235-243 (1986).

8   Hector, R. F. *et al*. The public health impact of coccidioidomycosis in Arizona and California. *Int J Environ Res Public Health* **8**, 1150-1173, doi:10.3390/ijerph8041150 (2011).

9   Barnato, A. E., Sanders, G. D. & Owens, D. K. Cost-effectiveness of a potential vaccine for Coccidioides immitis. *Emerging infectious diseases* **7**, 797-806, doi:10.3201/eid0705.010505 (2001).

10   Galgiani, J. N. *et al*. Practice guideline for the treatment of coccidioidomycosis. Infectious Diseases Society of America. *Clin Infect Dis* **30**, 658-661, doi:10.1086/313747 (2000).

11   Kim, M. M., Blair, J. E., Carey, E. J., Wu, Q. & Smilack, J. D. Coccidioidal pneumonia, Phoenix, Arizona, USA, 2000-2004. *Emerging infectious diseases* **15**, 397-401, doi:10.3201/eid1563.081007 (2009).

12   Grys, T. E. *et al*. Comparison of two FDA-cleared EIA assays for the detection of Coccidioides antibodies against a composite clinical standard. *Med Mycol*, doi:10.1093/mmy/myy094 (2018).

13   Malo, J. *et al*. Update on the diagnosis of pulmonary coccidioidomycosis. *Ann Am Thorac Soc* **11**, 243-253, doi:10.1513/AnnalsATS.201308-286FR (2014).

14   Kuberski, T. *et al*. Diagnosis of coccidioidomycosis by antigen detection using cross-reaction with a Histoplasma antigen. *Clin Infect Dis* **44**, e50-54, doi:10.1086/511684 (2007).

15   Kassis, C. *et al*. Role of Coccidioides Antigen Testing in the Cerebrospinal Fluid for the Diagnosis of Coccidioidal Meningitis. *Clin Infect Dis*, doi:10.1093/cid/civ585 (2015).

16   Rosenstein, N. E. *et al*. Risk factors for severe pulmonary and disseminated coccidioidomycosis: Kern County, California, 1995-1996. *Clin Infect Dis* **32**, 708-715, doi:10.1086/319203 (2001).

17   Nguyen, C. *et al*. Recent advances in our understanding of the environmental, epidemiological, immunological, and clinical dimensions of coccidioidomycosis. *Clin Microbiol Rev* **26**, 505-525, doi:10.1128/CMR.00005-13 (2013).

18   Wood, J. C., Friedly, G., Zartarian, M., Aarnaes, S. & de la Maza, L. M. Alternatives to the standardized laboratory branch complement fixation test for detection of antibodies to Coccidioides immitis. *J Clin Microbiol* **16**, 1030-1033 (1982).

19   Pappagianis, D. & Zimmer, B. L. Serology of coccidioidomycosis. *Clin Microbiol Rev* **3**, 247-268 (1990).

20   Kuberski, T., Herrig, J. & Pappagianis, D. False-positive IgM serology in coccidioidomycosis. *J Clin Microbiol* **48**, 2047-2049, doi:10.1128/JCM.01843-09 (2010).

21   Galgiani, J. N. *et al*. Coccidioidomycosis. *Clin Infect Dis* **41**, 1217-1223, doi:10.1086/496991 (2005).

22   Blair, J. E., Coakley, B., Santelli, A. C., Hentz, J. G. & Wengenack, N. L. Serologic testing for symptomatic coccidioidomycosis in immunocompetent and immunosuppressed hosts. *Mycopathologia* **162**, 317-324, doi:10.1007/s11046-006-0062-5 (2006).

23   Mendoza, N. & Blair, J. E. The utility of diagnostic testing for active coccidioidomycosis in solid organ transplant recipients. *Am J Transplant* **13**, 1034-1039, doi:10.1111/ajt.12144 (2013).

24   Borchers, A. T. & Gershwin, M. E. The immune response in Coccidioidomycosis. *Autoimmun Rev* **10**, 94-102, doi:10.1016/j.autrev.2010.08.010 (2010).

25   Durkin, M. *et al*. Diagnosis of coccidioidomycosis with use of the Coccidioides antigen enzyme immunoassay. *Clin Infect Dis* **47**, e69-73, doi:10.1086/592073 (2008).

26   Kirsch, E. J. *et al*. Evaluation of Coccidioides antigen detection in dogs with coccidioidomycosis. *Clin Vaccine Immunol* **19**, 343-345, doi:10.1128/CVI.05631-11 (2012).

27   Olson, E. J., Standing, J. E., Griego-Harper, N., Hoffman, O. A. & Limper, A. H. Fungal beta-glucan interacts with vitronectin and stimulates tumor necrosis factor alpha release from macrophages. *Infect Immun* **64**, 3548-3554 (1996).

28   Rop, O., Mlcek, J. & Jurikova, T. Beta-glucans in higher fungi and their health effects. *Nutr Rev* **67**, 624-631, doi:10.1111/j.1753-4887.2009.00230.x (2009).

29   Stevens, D. A. *et al*. Cerebrospinal Fluid (1,3)-Beta-d-Glucan Testing Is Useful in Diagnosis of Coccidioidal Meningitis. *J Clin Microbiol* **54**, 2707-2710, doi:10.1128/JCM.01224-16 (2016).

30   Thompson, G. R., 3rd *et al*. Serum (1->3)-beta-D-glucan measurement in coccidioidomycosis. *J Clin Microbiol* **50**, 3060-3062, doi:10.1128/JCM.00631-12 (2012).

31 Zangeneh, T. T. *et al*. Positive (1-3) B-d-glucan and cross reactivity of fungal assays in coccidioidomycosis. *Med Mycol* **53**, 171-173, doi:10.1093/mmy/myu077 (2015).

32 *Essentials of Glycobiology*. 3rd edn, (Cold Spring Harbor Laboratory Press, 2017).

33 Kruger, M. & Linke, W. A. The giant protein titin: a regulatory node that integrates myocyte signaling pathways. *J Biol Chem* **286**, 9905-9912, doi:10.1074/jbc.R110.173260 (2011).

34 George, J. & Sabapathi, S. N. Cellulose nanocrystals: synthesis, functional properties, and applications. *Nanotechnol Sci Appl* **8**, 45-54, doi:10.2147/NSA.S64386 (2015).

35 An, H. J., Froehlich, J. W. & Lebrilla, C. B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* **13**, 421-426, doi:10.1016/j.cbpa.2009.07.022 (2009).

36 Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).

37 Singh, V. P., Bali, A., Singh, N. & Jaggi, A. S. Advanced glycation end products and diabetic complications. *Korean J Physiol Pharmacol* **18**, 1-14, doi:10.4196/kjpp.2014.18.1.1 (2014).

38 Goettig, P. Effects of Glycosylation on the Enzymatic Activity and Mechanisms of Proteases. *Int J Mol Sci* **17**, doi:10.3390/ijms17121969 (2016).

39 Park, Y. J., Jeong, Y. U. & Kong, W. S. Genome Sequencing and Carbohydrate-Active Enzyme (CAZyme) Repertoire of the White Rot Fungus Flammulina elastica. *Int J Mol Sci* **19**, doi:10.3390/ijms19082379 (2018).

40 Henrissat, B. & Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* **7**, 637-644 (1997).

41 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495, doi:10.1093/nar/gkt1178 (2014).

42   Kapteyn, J. C., Dijkgraaf, G. J., Montijn, R. C. & Klis, F. M. Glucosylation of cell wall proteins in regenerating spheroplasts of Candida albicans. *FEMS Microbiol Lett* **128**, 271-277, doi:10.1111/j.1574-6968.1995.tb07535.x (1995).

43   Bernard, M. & Latge, J. P. Aspergillus fumigatus cell wall: composition and biosynthesis. *Med Mycol* **39 Suppl 1**, 9-17 (2001).

44   Kanetsuna, F., Carbonell, L. M., Gil, F. & Azuma, I. Chemical and ultrastructural studies on the cell walls of the yeastlike and mycelial forms of Histoplasma capsulatum. *Mycopathol Mycol Appl* **54**, 1-13 (1974).

45   Porter, J. F., Scheer, E. R. & Wheat, R. W. Characterization of 3-O-methylmannose from Coccidioides immitis. *Infect Immun* **4**, 660-661 (1971).

46   Wheat, R. W., Tritschler, C., Conant, N. F. & Lowe, E. P. Comparison of Coccidioides immitis arthrospore, mycelium, and spherule cell walls, and influence of growth medium on mycelial cell wall composition. *Infect Immun* **17**, 91-97 (1977).

47   Pappagianis, D., Putman, E. W. & Kobayashi, G. S. Polysaccharide of Coccidioides immitis. *J Bacteriol* **82**, 714-723 (1961).

48   Tarbet, J. E. & Breslau, A. M. Histochemical investigation of the spherule of Coccidioides immitis in relation to host reaction. *J Infect Dis* **92**, 183-190 (1953).

49   Szaniszlo, P. J. & Harris, J. L. *Fungal dimorphism : with emphasis on fungi pathogenic for humans*. (Plenum Press, 1985).

50   Lebrilla, C. B. & An, H. J. The prospects of glycan biomarkers for the diagnosis of diseases. *Mol Biosyst* **5**, 17-20, doi:10.1039/b811781k (2009).

51   Butvilovskaya, V. I. *et al*. [Modification of Anti-Glycan IgG and IgM Profiles in Allergic Inflammation]. *Mol Biol (Mosk)* **52**, 634-643, doi:10.1134/S0026898418040031 (2018).

52   Kamiyama, T. *et al*. Identification of novel serum biomarkers of hepatocellular carcinoma using glycomic analysis. *Hepatology* **57**, 2314-2325, doi:10.1002/hep.26262 (2013).

53   Nguyen-Khuong, T. *et al.* Glycomic characterization of basal tears and changes with diabetes and diabetic retinopathy. *Glycobiology* **25**, 269-283, doi:10.1093/glycob/cwu108 (2015).

54   Saldova, R. *et al.* Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology* **17**, 1344-1356, doi:10.1093/glycob/cwm100 (2007).

55   Thompson, S., Dargan, E., Griffiths, I. D., Kelly, C. A. & Turner, G. A. The glycosylation of haptoglobin in rheumatoid arthritis. *Clin Chim Acta* **220**, 107-114 (1993).

56   Paris, L. *et al.* Urine lipoarabinomannan glycan in HIV-negative patients with pulmonary tuberculosis correlates with disease severity. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aal2807 (2017).

57   Berg, J. M., Tymoczko, J. L., Stryer, L. & Stryer, L. *Biochemistry*. 5th edn, (W.H. Freeman, 2002).

58   Hu, S. & Wong, D. T. Lectin microarray. *Proteomics Clin Appl* **3**, 148-154, doi:10.1002/prca.200800153 (2009).

59   Matsuda, A. *et al.* Development of an all-in-one technology for glycan profiling targeting formalin-embedded tissue sections. *Biochem Biophys Res Commun* **370**, 259-263, doi:10.1016/j.bbrc.2008.03.090 (2008).

60   Haji-Ghassemi, O. *et al.* Molecular Basis for Recognition of the Cancer Glycobiomarker, LacdiNAc (GalNAc[beta1-->4]GlcNAc), by Wisteria floribunda Agglutinin. *J Biol Chem* **291**, 24085-24095, doi:10.1074/jbc.M116.750463 (2016).

61   West, C., Elfakir, C. & Lafosse, M. Porous graphitic carbon: a versatile stationary phase for liquid chromatography. *J Chromatogr A* **1217**, 3201-3216, doi:10.1016/j.chroma.2009.09.052 (2010).

62   Vial, J., Hennion, M. C., Fernandez-Alba, A. & Aguera, A. Use of porous graphitic carbon coupled with mass detection for the analysis of polar phenolic compounds by liquid chromatography. *J Chromatogr A* **937**, 21-29 (2001).

63   Liu, J. *et al*. Highly efficient N-glycoproteomic sample preparation by combining C(18) and graphitized carbon adsorbents. *Anal Bioanal Chem* **406**, 3103-3109, doi:10.1007/s00216-014-7716-9 (2014).

64   Grys, T. E. *et al*. Total and Lectin-Binding Proteome of Spherulin from Coccidioides posadasii. *J Proteome Res* **15**, 3463-3472, doi:10.1021/acs.jproteome.5b01054 (2016).

65   Latge, J. P. The cell wall: a carbohydrate armour for the fungal cell. *Mol Microbiol* **66**, 279-290, doi:10.1111/j.1365-2958.2007.05872.x (2007).

66   Hykollari, A. *et al*. More Than Just Oligomannose: An N-glycomic Comparison of Penicillium Species. *Mol Cell Proteomics* **15**, 73-92, doi:10.1074/mcp.M115.055061 (2016).

67   Deshpande, N., Wilkins, M. R., Packer, N. & Nevalainen, H. Protein glycosylation pathways in filamentous fungi. *Glycobiology* **18**, 626-637, doi:10.1093/glycob/cwn044 (2008).

68   Ospina Alvarez, S. P. *et al*. Comparison of extraction methods of chitin from Ganoderma lucidum mushroom obtained in submerged culture. *Biomed Res Int* **2014**, 169071, doi:10.1155/2014/169071 (2014).

69   Tintelnot, K. *et al*. Taxonomic and diagnostic markers for identification of Coccidioides immitis and Coccidioides posadasii. *Med Mycol* **45**, 385-393, doi:10.1080/13693780701288070 (2007).

70   Fisher, M. C. *et al*. Biogeographic range expansion into South America by Coccidioides immitis mirrors New World patterns of human migration. *Proc Natl Acad Sci U S A* **98**, 4558-4562, doi:10.1073/pnas.071406098 (2001).

71   Neafsey, D. E. *et al*. Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control. *Genome Res* **20**, 938-946, doi:10.1101/gr.103911.109 (2010).

72   Whiston, E. *et al*. Comparative transcriptomics of the saprobic and parasitic growth phases in Coccidioides spp. *PLoS One* **7**, e41034, doi:10.1371/journal.pone.0041034 (2012).

73   Castellana, N. & Bafna, V. Proteogenomics to discover the full coding content of genomes: a computational perspective. *J Proteomics* **73**, 2124-2135, doi:10.1016/j.jprot.2010.06.007 (2010).

74   Chen, G. *et al*. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* **1**, 304-313 (2002).

75   Nie, L., Wu, G., Culley, D. E., Scholten, J. C. & Zhang, W. Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Crit Rev Biotechnol* **27**, 63-75, doi:10.1080/07388550701334212 (2007).

76   Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-1730 (1999).

77   Cox, R. A. & Britt, L. A. Isolation of a coccidioidin component that reacts with immunoglobulin M precipitin antibody. *Infect Immun* **53**, 449-453 (1986).

78   Cox, R. A. & Britt, L. A. Antigenic identity of biologically active antigens in coccidioidin and spherulin. *Infect Immun* **55**, 2590-2596 (1987).

79   Taylor, C. F. *et al*. The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol* **25**, 887-893, doi:10.1038/nbt1329 (2007).

80   Vizcaino, J. A. *et al*. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* **41**, D1063-1069, doi:10.1093/nar/gks1262 (2013).

81   Sharpton, T. J. *et al*. Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. *Genome Res* **19**, 1722-1731, doi:10.1101/gr.087551.108 (2009).

82   Stajich, J. E. *et al*. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res* **40**, D675-681, doi:10.1093/nar/gkr918 (2012).

83   Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**, W435-439, doi:10.1093/nar/gkl200 (2006).

84   Henn, M. R. *et al*.   (2007).

85   Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534-2536, doi:10.1093/bioinformatics/btn323 (2008).

86   Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **5**, 5277, doi:10.1038/ncomms6277 (2014).

87   Ma, Z. Q. *et al*. IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J Proteome Res* **8**, 3872-3881, doi:10.1021/pr900360j (2009).

88   Robinson, J. T. *et al*. Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

89   Coordinators, N. R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **44**, D7-19, doi:10.1093/nar/gkv1290 (2016).

90   Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265-268, doi:10.1093/nar/gkm286 (2007).

91   Owens, R. A., Hammel, S., Sheridan, K. J., Jones, G. W. & Doyle, S. A proteomic approach to investigating gene cluster expression and secondary metabolite functionality in Aspergillus fumigatus. *PLoS One* **9**, e106942, doi:10.1371/journal.pone.0106942 (2014).

92   Chang, S. S., Zhang, Z. & Liu, Y. RNA interference pathways in fungi: mechanisms and functions. *Annu Rev Microbiol* **66**, 305-323, doi:10.1146/annurev-micro-092611-150138 (2012).

93   Nicolas, F. E. & Ruiz-Vazquez, R. M. Functional diversity of RNAi-associated sRNAs in fungi. *Int J Mol Sci* **14**, 15348-15360, doi:10.3390/ijms140815348 (2013).

94   Nicolas, F. E. & Garre, V. RNA Interference in Fungi: Retention and Loss. *Microbiol Spectr* **4**, doi:10.1128/microbiolspec.FUNK-0008-2016 (2016).

95   Muszewska, A., Hoffman-Sommer, M. & Grynberg, M. LTR retrotransposons in fungi. *PLoS One* **6**, e29425, doi:10.1371/journal.pone.0029425 (2011).

96   Maringer, K. *et al*. Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in Aedes aegypti. *BMC Genomics* **18**, 101, doi:10.1186/s12864-016-3432-5 (2017).

97   Schwenk, J. M. *et al*. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J Proteome Res* **16**, 4299-4310, doi:10.1021/acs.jproteome.7b00467 (2017).

98   Orsborn, K. I. *et al*. Protein expression profiling of Coccidioides posadasii by two-dimensional differential in-gel electrophoresis and evaluation of a newly recognized peroxisomal matrix protein as a recombinant vaccine candidate. *Infect Immun* **74**, 1865-1872, doi:10.1128/IAI.74.3.1865-1872.2006 (2006).

99   Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845-867 (2002).

100   Schiess, R., Wollscheid, B. & Aebersold, R. Targeted proteomic strategy for clinical biomarker discovery. *Mol Oncol* **3**, 33-44, doi:10.1016/j.molonc.2008.12.001 (2009).

101   Becker, J. M. *et al*. Pathway analysis of Candida albicans survival and virulence determinants in a murine infection model. *Proc Natl Acad Sci U S A* **107**, 22044-22049, doi:10.1073/pnas.1009845107 (2010).

102   Parente-Rocha, J. A. *et al*. In vitro, ex vivo and in vivo models: A comparative analysis of Paracoccidioides spp. proteomic studies. *Fungal Biol* **122**, 505-513, doi:10.1016/j.funbio.2017.10.009 (2018).

103   Mead, H. L., Teixeira, M. M., Galgiani, J. N. & Barker, B. M. Characterizing in vitro spherule morphogenesis of multiple strains of both species of Coccidioides. *Med Mycol*, doi:10.1093/mmy/myy049 (2018).

104   Levine, H. B., Cobb, J. M. & Smith, C. E. Immunogenicity of spherule-endospore vaccines of Coccidioides immitis for mice. *J Immunol* **87**, 218-227 (1961).

105   Jensen, P. H., Karlsson, N. G., Kolarich, D. & Packer, N. H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* **7**, 1299-1310, doi:10.1038/nprot.2012.063 (2012).

106   Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, doi:10.1371/journal.pone.0021800 (2011).

107   Basenko, E. Y. *et al*. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. *J Fungi (Basel)* **4**, doi:10.3390/jof4010039 (2018).

108   Kolde, R. & Vilo, J. GOsummaries: an R Package for Visual Functional Annotation of Experimental Data. *F1000Res* **4**, 574, doi:10.12688/f1000research.6925.1 (2015).

109   Steentoft, C. *et al*. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **32**, 1478-1488, doi:10.1038/emboj.2013.79 (2013).

110   Chuang, G. Y. *et al*. Computational prediction of N-linked glycosylation incorporating structural properties and patterns. *Bioinformatics* **28**, 2249-2255, doi:10.1093/bioinformatics/bts426 (2012).

111   Florens, L. *et al*. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40**, 303-311, doi:10.1016/j.ymeth.2006.07.028 (2006).

112   Lewis, E. R. *et al*. Differences in Host Innate Responses among Coccidioides Isolates in a Murine Model of Pulmonary Coccidioidomycosis. *Eukaryot Cell* **14**, 1043-1053, doi:10.1128/EC.00122-15 (2015).

113   Hung, C. Y., Yu, J. J., Seshan, K. R., Reichard, U. & Cole, G. T. A parasitic phase-specific adhesin of Coccidioides immitis contributes to the virulence of this respiratory Fungal pathogen. *Infect Immun* **70**, 3443-3456 (2002).

114   Champer, J. *et al*. Protein targets for broad-spectrum mycosis vaccines: quantitative proteomic analysis of Aspergillus and Coccidioides and comparisons with other fungal pathogens. *Ann N Y Acad Sci* **1273**, 44-51, doi:10.1111/j.1749-6632.2012.06761.x (2012).

115   Galgiani, J. N., Hayden, R. & Payne, C. M. Leukocyte effects on the dimorphism of Coccidioides immitis. *J Infect Dis* **146**, 56-63 (1982).

116   Converse, J. L. Effect of surface active agents on endosporulation of Coccidioides immitis in a chemically defined medium. *J Bacteriol* **74**, 106-107 (1957).

117   Bao, Z. *et al.* Evidence for the involvement of cofilin in Aspergillus fumigatus internalization into type II alveolar epithelial cells. *BMC Microbiol* **15**, 161, doi:10.1186/s12866-015-0500-y (2015).

118   Karkowska-Kuleta, J. & Kozik, A. Cell wall proteome of pathogenic fungi. *Acta Biochim Pol* **62**, 339-351, doi:10.18388/abp.2015_1032 (2015).

119   Skropeta, D. The effect of individual N-glycans on enzyme activity. *Bioorg Med Chem* **17**, 2645-2653, doi:10.1016/j.bmc.2009.02.037 (2009).

120   Huang, L. *et al.* dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res* **46**, D516-D521, doi:10.1093/nar/gkx894 (2018).

121   Varki, A. & Lowe, J. B. in *Essentials of Glycobiology* (eds nd *et al.*) (2009).

122   Nightingale, A. *et al.* The Proteins API: accessing key integrated protein and genome information. *Nucleic Acids Res* **45**, W539-W544, doi:10.1093/nar/gkx237 (2017).

123   Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**, W445-451, doi:10.1093/nar/gks479 (2012).

124   Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**, W95-W101, doi:10.1093/nar/gky418 (2018).

125   Mitchell, N. M. *et al.* Proteogenomic Re-Annotation of Coccidioides posadasii Strain Silveira. *Proteomics* **18**, doi:10.1002/pmic.201700173 (2018).

126   Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5-9, doi:10.1093/nar/gkn201 (2008).

127   Miyazaki, T., Sugisawa, T. & Hoshino, T. Pyrroloquinoline quinone-dependent dehydrogenases from Ketogulonicigenium vulgare catalyze the direct conversion of L-sorbosone to L-ascorbic acid. *Appl Environ Microbiol* **72**, 1487-1495, doi:10.1128/AEM.72.2.1487-1495.2006 (2006).

128   Kirkland, T. N. & Fierer, J. Coccidioides immitis and posadasii; A review of their biology, genomics, pathogenesis, and host immunity. *Virulence* **9**, 1426-1435, doi:10.1080/21505594.2018.1509667 (2018).

129   Carapito, R. *et al*. Molecular basis of arabinobio-hydrolase activity in phytopathogenic fungi: crystal structure and catalytic mechanism of Fusarium graminearum GH93 exo-alpha-L-arabinanase. *J Biol Chem* **284**, 12285-12296, doi:10.1074/jbc.M900439200 (2009).

130   Martens-Uzunova, E. S. & Schaap, P. J. Assessment of the pectin degrading enzyme network of Aspergillus niger by functional genomics. *Fungal Genet Biol* **46 Suppl 1**, S170-S179 (2009).

131   Sprockett, D. D., Piontkivska, H. & Blackwood, C. B. Evolutionary analysis of glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions in necrotrophic fungal pathogens. *Gene* **479**, 29-36, doi:10.1016/j.gene.2011.02.009 (2011).

132   Seymour, G. B. & Knox, J. P. *Pectins and their manipulation*. (Blackwell ; Published in U.S. and Canada only by CRC Press, 2002).

133   Teixeira, M. M. *et al*. Comparative genomics of the major fungal agents of human and animal Sporotrichosis: Sporothrix schenckii and Sporothrix brasiliensis. *BMC Genomics* **15**, 943, doi:10.1186/1471-2164-15-943 (2014).

134   Hung, C. Y., Yu, J. J., Lehmann, P. F. & Cole, G. T. Cloning and expression of the gene which encodes a tube precipitin antigen and wall-associated beta-glucosidase of Coccidioides immitis. *Infect Immun* **69**, 2211-2222, doi:10.1128/IAI.69.4.2211-2222.2001 (2001).

135   Medeiros, A. I. *et al*. Leukotrienes are potent adjuvant during fungal infection: effects on memory T cells. *J Immunol* **181**, 8544-8551 (2008).

136   Caffrey-Carr, A. K. *et al*. Host-Derived Leukotriene B4 Is Critical for Resistance against Invasive Pulmonary Aspergillosis. *Front Immunol* **8**, 1984, doi:10.3389/fimmu.2017.01984 (2017).

137   Noverr, M. C., Toews, G. B. & Huffnagle, G. B. Production of prostaglandins and leukotrienes by pathogenic fungi. *Infect Immun* **70**, 400-402 (2002).

138   Duhring, S. *et al*. Host-pathogen interactions between the human innate immune system and Candida albicans-understanding and modeling defense and evasion strategies. *Front Microbiol* **6**, 625, doi:10.3389/fmicb.2015.00625 (2015).

139   Ene, I. V., Cheng, S. C., Netea, M. G. & Brown, A. J. Growth of Candida albicans cells on the physiologically relevant carbon source lactate affects their recognition and phagocytosis by immune cells. *Infect Immun* **81**, 238-248, doi:10.1128/IAI.01092-12 (2013).

140   Tiemeyer, M. *et al*. GlyTouCan: an accessible glycan structure repository. *Glycobiology* **27**, 915-919, doi:10.1093/glycob/cwx066 (2017).

141   New, J. S., King, R. G. & Kearney, J. F. Manipulation of the glycan-specific natural antibody repertoire for immunotherapy. *Immunol Rev* **270**, 32-50, doi:10.1111/imr.12397 (2016).

142   Muszewska, A., Pilsyk, S., Perlinska-Lenart, U. & Kruszewska, J. S. Diversity of Cell Wall Related Proteins in Human Pathogenic Fungi. *J Fungi (Basel)* **4**, doi:10.3390/jof4010006 (2017).

143   Watanabe, S., Saimura, M. & Makino, K. Eukaryotic and bacterial gene clusters related to an alternative pathway of nonphosphorylated L-rhamnose metabolism. *J Biol Chem* **283**, 20372-20382, doi:10.1074/jbc.M801065200 (2008).

144   Couturier, M. R., Graf, E. H. & Griffin, A. T. Urine antigen tests for the diagnosis of respiratory infections: legionellosis, histoplasmosis, pneumococcal pneumonia. *Clin Lab Med* **34**, 219-236, doi:10.1016/j.cll.2014.02.002 (2014).

145   Pattathil, S. *et al*. A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol* **153**, 514-525, doi:10.1104/pp.109.151985 (2010).

146   Grun, C. H. *et al*. One-step biotinylation procedure for carbohydrates to study carbohydrate-protein interactions. *Anal Biochem* **354**, 54-63, doi:10.1016/j.ab.2006.03.055 (2006).

147   Kalay, H., Ambrosini, M., Chiodo, F., van Kooyk, Y. & Garcia-Vallejo, J. J. Enhanced glycan nanoprofiling by weak anion exchange preparative chromatography, mild acid desialylation, and nanoliquid chromatography-mass

spectrometry with nanofluorescence detection. *Electrophoresis* **34**, 2350-2356, doi:10.1002/elps.201200657 (2013).

148   Rohrlich, P. *et al*. Prospective sandwich enzyme-linked immunosorbent assay for serum galactomannan: early predictive value and clinical use in invasive aspergillosis. *Pediatr Infect Dis J* **15**, 232-237 (1996).

149   Burton, E. N. *et al*. Characterization of the urinary microbiome in healthy dogs. *PLoS One* **12**, e0177783, doi:10.1371/journal.pone.0177783 (2017).

150   Lee, W. *et al*. Discrimination of Human Urine from Animal Urine Using 1H-NMR. *J Anal Toxicol* **43**, 51-60, doi:10.1093/jat/bky061 (2019).

151   Zhou, S., Wooding, K. M. & Mechref, Y. Analysis of Permethylated Glycan by Liquid Chromatography (LC) and Mass Spectrometry (MS). *Methods Mol Biol* **1503**, 83-96, doi:10.1007/978-1-4939-6493-2_7 (2017).

152   Kailemia, M. J., Ruhaak, L. R., Lebrilla, C. B. & Amster, I. J. Oligosaccharide analysis by mass spectrometry: a review of recent developments. *Anal Chem* **86**, 196-212, doi:10.1021/ac403969n (2014).

153   Valdivia, L. *et al*. Coccidioidomycosis as a common cause of community-acquired pneumonia. *Emerg Infect Dis* **12**, 958-962 (2006).

154   Chang, D. C. *et al*. Testing for coccidioidomycosis among patients with community-acquired pneumonia. *Emerg Infect Dis* **14**, 1053-1059, doi:10.3201/eid1407.070832 (2008).

155   Bezold, C. P. *et al*. Notes from the Field: Increase in Coccidioidomycosis - Arizona, October 2017-March 2018. *MMWR Morb Mortal Wkly Rep* **67**, 1246-1247, doi:10.15585/mmwr.mm6744a6 (2018).

156   Ampel, N. M. The diagnosis of coccidioidomycosis. *F1000 Med Rep* **2**, doi:10.3410/M2-2 (2010).

157   Richer, S. M. *et al*. Improved Diagnosis of Acute Pulmonary Histoplasmosis by Combining Antigen and Antibody Detection. *Clin Infect Dis* **62**, 896-902, doi:10.1093/cid/ciw007 (2016).

158   Haji-Ghassemi, O., Blackler, R. J., Martin Young, N. & Evans, S. V. Antibody recognition of carbohydrate epitopesdagger. *Glycobiology* **25**, 920-952, doi:10.1093/glycob/cwv037 (2015).

159   Marino, C., Rinflerch, A. & de Lederkremer, R. M. Galactofuranose antigens, a target for diagnosis of fungal infections in humans. *Future Sci OA* **3**, FSO199, doi:10.4155/fsoa-2017-0030 (2017).

160   Pinto, M. R., Barreto-Bergter, E. & Taborda, C. P. Glycoconjugates and polysaccharides of fungal cell wall and activation of immune system. *Braz J Microbiol* **39**, 195-208, doi:10.1590/S1517-83822008000200001 (2008).

161   Turner, A. E. B. *et al.* Novel polysaccharide-protein conjugates provide an immunogenic 13-valent pneumococcal conjugate vaccine for S. pneumoniae. *Synth Syst Biotechnol* **2**, 49-58, doi:10.1016/j.synbio.2016.12.002 (2017).

APPENDIX A

STATEMENT OF PERMISSIONS

All co-authors have granted their permission for the use of the articles presented in this dissertation.

APPENDIX B

LIST OF 288 NOVEL PEPTIDES

	Peptide	Notes	AUG	Six	EST	Peptide Classification	Novel Gene Model/ locati...
1	EGLKPTIFVVNNAGYSIER		No	No	Yes	Peptides crossing intron-exon boundaries	g1825.t1
2	LIHGPMQQYNDISTQWDYQK		No	No	Yes	Peptides crossing intron-exon boundaries	g1825.t1
3	TVTTTAVSSMGIIPQNFHLENPSR		No	No	Yes	Peptides crossing intron-exon boundaries	g1061.t1
4	VILLVGDGSLQMTVQEFGTIIR		No	No	Yes	Peptides crossing intron-exon boundaries	g1825.t1
5	VLMVMYDGGEHAK		No	No	Yes	Peptides crossing intron-exon boundaries	g366.t1
6	ANLDYQR	Contig extension, contigs 69-70 (in gap)	No	No	Yes	Peptides outside of annotated contig	c69-70
7	DFSTAAEFELDLVNVK	Contig extension, contig 148-149 (partially in gap)	No	No	Yes	Peptides outside of annotated contig	c148-149
8	DIGYDDSSK	Contig extension, contigs 69-70 (in gap)	No	No	Yes	Peptides outside of annotated contig	c69-70
9	DMEGVDAVPGAASR	Contig gap protein - XP_001246931.2	No	No	Yes	Peptides outside of annotated contig	cUnknown1
10	FADTEEIPLPPFWGGVR	Contig gap protein - XP_001246931.2	No	No	Yes	Peptides outside of annotated contig	cUnknown1
11	LDGSLPWLRPDTK	Contig extension, contigs 69-70	No	No	Yes	Peptides outside of annotated contig	c69-70
12	LGAGDQGIMFGYATDETPELLPLTIQLAHK	Contig extension, contigs 69-70 (in gap)	No	No	Yes	Peptides outside of annotated contig	c69-70
13	LLPSDPAELFTSIPSRPFFPR	Contig extension, contigs 164-165 (in gap, Glucosidase I)	No	No	Yes	Peptides outside of annotated contig	c164-165
14	LVPESVEFWQGR	Contig gap protein - XP_001246931.2	No	No	Yes	Peptides outside of annotated contig	cUnknown1
15	SGASTPSTNAAVVSR	Contig gap protein - XP_001246931.2	No	No	Yes	Peptides outside of annotated contig	cUnknown1
16	TCNVLVAIEQQSPDIAQGLHYDEALEK	Contig extension, contigs 69-70 (in gap)	No	No	Yes	Peptides outside of annotated contig	c69-70
17	TGMIMVFGEITTK	Contig extension, contigs 69-70 (in gap)	No	No	Yes	Peptides outside of annotated contig	c69-70
18	TQVTVEYAHDNGAMKPLR	Contig extension, contigs 69-70	No	No	Yes	Peptides outside of annotated contig	c69-70
19	VDTVVVSAQHSEDITTEELRK	Contig extension, contigs 69-70	No	No	Yes	Peptides outside of annotated contig	c69-70
20	VENYLTTELK	Contig extension, contig 148-149 (in gap)	No	No	Yes	Peptides outside of annotated contig	c148-149
21	VHGENEHLGEGDEK	Contig gap protein - XP_001246931.2	No	No	Yes	Peptides outside of annotated contig	cUnknown1
22	VIPENLLDDR	Contig extension, contigs 69-70	No	No	Yes	Peptides outside of annotated contig	c69-70
23	AACAPVQAQVDMSVYYEQVYR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_04847
24	AFENTQGPGSPFQGVGEGSWENGVWDYK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_05111
25	AGKNADNTPLLMFALPHHVEAFDDSSK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_01467
26	ASISVSDELFPFSR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_01467
27	ASSMSIPEAPAEGGFRSVVYFVNWAIYGR	SNP, Crosses splice boundary	No	No	Yes	Single nucleotide polymorphisms	CIST_07834
28	AVGSAVTLGAGVQLSELYK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIHT_02754
29	DCTLQACGAEDLPK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_07975
30	DGILGDVAPTILDAMGIEQPK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CPAT_00917
31	DVFEPSQLAEKMMTEEDNEIRFTDEPER	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_02400
32	EKDGILGDVAPTILDAMGIEQPK	SNP	No	No	Yes	Single nucleotide polymorphisms	CPAT_00917
33	ESNIQAPIYPTK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIMG_08465T0
34	ETDYTMGEWAANEQGGIR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_08556
35	ETGGTRDVGYAVR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_07880
36	EVLQEESNVQPVK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIMG_09663T0
37	FADHVSAFPAGINVGATWSKNLAYLR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_02008
38	FLELYAPQLKGTTFDVVSINGGR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_02010
39	GENIIFRPSVNWK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CPAT_05740
40	GFFDNDELTR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_02653
41	GHAPGPSLDTTSVGENIIFR	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CPAT_05740
42	GPDTDLGVLNGLLK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIHT_01037
43	GPLHGLPILIK	SNP, Crosses splice boundary	No	No	Yes	Single nucleotide polymorphisms	CIRT_02549
44	GQSVDPISLLEKLLPLYEQLLIQLK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_06617
45	GSYLVTDHIMSELPEIR	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIMG_06416T0
46	GTVTFEQADENSPTTISWNISGHDANAQR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_01113
47	GWDVDSTEGK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_04788
48	HHGIMIGMGQK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIMG_02791T0
49	HVQNTLFQCIGGPNGEIKIIK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_05743
50	IGIESYFCSQDCFK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIHT_01770

168

	Peptide	Notes	AUG	Six	EST	Peptide Classification	Novel Gene Model/ locati
51	IPEAPAEGGFRSVVYFVNWAIYGR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_07834
52	ISCTSTDQLTTFTLDVNSELYR	SNP / Insertion	No	No	Yes	Single nucleotide polymorphisms	CIRT_00702
53	ISGTGSVMSASPLAATTITNVQTR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_07115
54	IVGKVTDDEIHWYANQVVEGAPPMK	SNP	No	No	Yes	Single nucleotide polymorphisms	CPAT_08282
55	IVGKVTDNEIHWYANQVVEGAPPMK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_07115
56	IYHVHDQPVPENGDCLGTLAHLDPYER	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_09572
57	KMGGGTLSKLVTDTLWAAGK	SNP	No	No	Yes	Single nucleotide polymorphisms	gi|303315515|ref|XP_003
58	KVLDSAKPDQIVYFDHGAYLITNTVK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_00218
59	KVSGHMSNVFSSTTKPESTPFSSDK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
60	KYMASTQMEATDAR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_01989
61	LGAVASESSICSGFGIDMLK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_09416
62	LGLNSGHLTCFDGGNFLLGGQILDR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_00356
63	LGLNSGHLTCFDGGNFLLGGQILDRDDFTK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_00356
64	LIPHACIQAAGADSR	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIHT_02076
65	LNIGVAVSADPNGHMTTDSVSTALLQR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_01062
66	LVQASSMSSMPNSYPVPEAPAEGGFR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
67	NADNTPLLMFALPHHVE	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_01467
68	NADNTPLLMFALPHHVEAFDDSSK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_01467
69	NADNTPLLMFALPHHVEAFDDSSKR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_01467
70	NSYPVPEAPAEGGFRSVVYFVNWAIYGR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
71	QAAVTDEMHNRQSQFPER	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_06617
72	QDLEMAGVMGIDPETVFADR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_04847
73	QTLVQASSMSSMPNSYPVPEAPAEGGFR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
74	RPSELQQELQVDFPDSGQGADGLLQILDR	SNP	No	No	Yes	Single nucleotide polymorphisms	gi|303316970|ref|XP_003
75	SGIPTFVAPMGK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	gi|597574547|ref|XP_007
76	SGTSMAAAHVSGAGAYIMAIEK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIHT_08660
77	SLSQIESISNNEGSVGDFLVEYLER	SNP	No	No	Yes	Single nucleotide polymorphisms	gi|303313822|ref|XP_003
78	SMPNSYPVPEAPAEGGFR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
79	SMSSMPNSYPVPEAPAEGGFR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
80	SNIQAPIYPTK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIMG_08465T0
81	SSMPNSYPVPEAPAEGGFR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
82	SYPVPEAPAEGGFR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
83	**TGNESLVGTVVNGFGGTGKLEQR**	**Endochitinase variant (GFGG)**	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
84	**TGNESLVGTVVNGFGGTGKLEQRE**	**Endochitinase variant (GFGG)**	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
85	TKLQPLNYLYNK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_01085
86	TLAIHPWSTTHEQLSDEEKG	SNP	No	No	Yes	Single nucleotide polymorphisms	gi|597569899|ref|XP_007
87	TVVVMHTAGPVLVDK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIST_02008
88	VCAETTFLPFDDTNWDALSNALDQCR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIMG_00948T0
89	VQASSMSSMPNSYPVPEAPAEGGFR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
90	VRPSDDPFSIITPSDPEAR	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_07726
91	VSGAGAYIMAIEK	SNP / Leucine-Isoleucine	No	No	Yes	Single nucleotide polymorphisms	CIHT_08660
92	VSGHMSNVFSSTTKPESTPFSSDK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIRT_10093
93	VTDNEIHWYANQVVEGAPPMK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_07115
94	VTDQTELEGLMDEAAYK	SNP	No	No	Yes	Single nucleotide polymorphisms	CIHT_05234
95	YLSDTWADTDKHYPGDKWDEPGKNVYGCTK	SNP	No	No	Yes	Single nucleotide polymorphisms	gi|303322170|ref|XP_003
96	AYLGVAADLSHINTNSTVTGHDPTPSGLR	Boundary adjustment, g4178.t1	No	Yes	No	Peptides crossing intron-exon boundaries	g4178.t1
97	QGDWILFTHEGGVDVGDVDAK	Boundary adjustment, g2104.t1	No	Yes	No	Peptides crossing intron-exon boundaries	g2104.t1
98	**SHRWMQEACVGNVGSIPR**	**Multiple boundary adjustment, g2406.t1**	No	Yes	No	Peptides crossing intron-exon boundaries	g2406.t1
99	VIQDLATAPGPYQK	Boundary adjustment, g3248.t1	No	Yes	No	Peptides crossing intron-exon boundaries	g3248.t1
100	WIGAFTPTCSASHLPGYIQK	Boundary adjustment/frame adjustment, g3367.t1	No	Yes	No	Peptides crossing intron-exon boundaries	g3367.t1

169

	Peptide	Notes	AUG	Six	EST	Peptide Classification	Novel Gene Model/ locati
101	AAEMPVDTSQR	Exon boundary adjustment, g646.t1	No	Yes	Yes	Peptides crossing intron-exon boundaries	g646.t1
102	AEMPVDTSQR	Exon boundary adjustment, g646.t1	No	Yes	Yes	Peptides crossing intron-exon boundaries	g646.t1
103	AINFGVAIFDDLDITR	Exon boundary adjustment, g5249.t1	No	Yes	Yes	Peptides crossing intron-exon boundaries	g5249.t1
104	IFDDLDITR	Exon boundary adjustment, g5249.t1	No	Yes	Yes	Peptides crossing intron-exon boundaries	g5249.t1
105	MANSMPNGCQDQVAVCR	Exon boundary adjustment, g4405.t1	No	Yes	Yes	Peptides crossing intron-exon boundaries	g4405.t1
106	SYCSGAQSMCR	Exon boundary adjustment, g4405.t1	No	Yes	Yes	Peptides crossing intron-exon boundaries	g4405.t1
107	TSLVPYSYCSGAQSMCR	Exon boundary adjustment, g4405.t1	No	Yes	Yes	Peptides crossing intron-exon boundaries	g4405.t1
108	EEKDATESIVNR	5' UTR, Contig 290	No	Yes	No	Peptides mapped to 5' UTRs	5p1
109	NLEEKDATESIVNR	5' UTR, Contig 290	No	Yes	No	Peptides mapped to 5' UTRs	5p1
110	AEMPELPTVPK		Yes	Yes	Yes	Novel exon	g417.t1
111	CGVSSLVLK		Yes	Yes	Yes	Novel exon	g1162.t1
112	DLPHGIPTELRPK		Yes	Yes	Yes	Novel exon	g3647.t1
113	EAVWGADPLEAHGGFR		Yes	Yes	Yes	Novel exon	g543.t1
114	ERAEMPELPTVPK		Yes	Yes	Yes	Novel exon	g417.t1
115	EWVAEWVEDALTLGVGVIAQR		Yes	Yes	Yes	Novel exon	g4535.t1
116	FAEATNEPGLAFAFGR		Yes	Yes	Yes	Novel exon	g1628.t1
117	FDGILGLGYDTISVNK		Yes	Yes	Yes	Novel exon	g1628.t1
118	GFVSQDTLR		Yes	Yes	Yes	Novel exon	g1628.t1
119	GILGLGYDTISVNK		Yes	Yes	Yes	Novel exon	g1628.t1
120	GMPAEFSLSTVHLDPLTR		Yes	Yes	Yes	Novel exon	g6519.t1
121	GQDFAEATNEPGLAFAFGR		Yes	Yes	Yes	Novel exon	g1628.t1
122	GSYLVTDHIMSELPELR		Yes	Yes	Yes	Novel exon	g471.t1
123	IGDLTIEGQDFAEATNEPGLAFAFGR		Yes	Yes	Yes	Novel exon	g1628.t1
124	IVPPFYNMINEGLIDEPVFGFYLGDTNK		Yes	Yes	Yes	Novel exon	g1628.t1
125	LSVGTEHIDDIIADFEQSFAVAADSSK		Yes	Yes	Yes	Novel exon	g554.t1
126	LSVGTEHIDDIIADFEQSFAVAADSSKK		Yes	Yes	Yes	Novel exon	g554.t1
127	MGVGEGGVGK		Yes	Yes	Yes	Novel exon	g4535.t1
128	MKVEQATVVESGAGEAAR		Yes	Yes	Yes	Novel exon	g4535.t1
129	NEPGLAFAFGR		Yes	Yes	Yes	Novel exon	g1628.t1
130	PAPWRPLFQANCPK		Yes	Yes	Yes	Novel exon	g6519.t1
131	PLSVVAVVCNNR		Yes	Yes	Yes	Novel exon	g6812.t1
132	SLDLPHGIPTELRPK		Yes	Yes	Yes	Novel exon	g3647.t1
133	SVVAVVCNNR		Yes	Yes	Yes	Novel exon	g6812.t1
134	TYAELGK	Part of pyruvate decarboxylase	Yes	Yes	Yes	Novel exon	g1825.t1
135	VEQATVVESGAGEAAR		Yes	Yes	Yes	Novel exon	g4535.t1
136	VLNDEAFK	Part of pyruvate decarboxylase	Yes	Yes	Yes	Novel exon	g1825.t1
137	VVAVVCNNR		Yes	Yes	Yes	Novel exon	g6812.t1
138	VVLDTGSSNLWVPSSECGSIACYLH		Yes	Yes	Yes	Novel exon	g1628.t1
139	VVLDTGSSNLWVPSSECGSIACYLHNK		Yes	Yes	Yes	Novel exon	g1628.t1
140	YDSSASSTYK		Yes	Yes	Yes	Novel exon	g1628.t1
141	YDSSASSTYKK		Yes	Six	Yes	Novel exon	g1628.t1
142	YGIKPLSVVAVVCNNR		Yes	Yes	Yes	Novel exon	g6812.t1
143	YGSGSLSGFVSQDTLR		Yes	Yes	Yes	Novel exon	g1628.t1
144	YVIVTLGSEGVWYSAADVGSK		Yes	Yes	Yes	Novel exon	g441.t1
145	FQIGSLTK		No	No	Yes	Novel exon	CIRT_01518
146	IDYVYLLIK		No	No	Yes	Novel exon	CIRT_00702
147	MGHVDHGK		No	No	Yes	Novel exon	CIRT_07923
148	MGHVDHGKTTILDWLRKSSV		No	No	Yes	Novel exon	CIRT_07923
149	SDTWADTDKHYPGDKWDEPGNNVYGCIK		No	No	Yes	Novel exon	CIRT_10093
150	TLNDGTVLEEVVVDAPLGHR		No	No	Yes	Novel exon	CPAT_07711

	Peptide	Notes	AUG	Six	EST	Peptide Classification	Novel Gene Model/ locati
151	TRDWAVESELREAAR		No	No	Yes	Novel exon	CIRT_01518
152	LYEIGAGTSEIR		No	Yes	No	Novel exon	cont275.1.153
153	SGGTTMYPGIADR		No	Yes	No	Novel exon	cont330.1.8
154	**AAAHIGIPVAMVNLGFASWIQTLCK**		Yes	No	No	Novel exon	g2026.t1
155	ADRLSVADGNRLSSSR		Yes	No	No	Novel exon	g882.t1
156	ALQVTLNQQGGR		Yes	No	No	Novel exon	g3677.t1
157	ARFSRSKPKVPQTIDSPWHCRLLPRFPMR		Yes	No	No	Novel exon	g3129.t1
158	ATGQQPPPAAEAVQEER		Yes	No	No	Novel exon	g6058.t1
159	AVADAFLFTVCDK		Yes	No	No	Novel exon	g7188.t1
160	AVLNPYANVDIR		Yes	No	No	Novel exon	g5554.t1
161	AYFSNYGKCTDIFAPGLNILSTWIGSK		Yes	No	No	Novel exon	g5092.t1
162	AYSMPSDSERETFVEADVKGGTNAM		Yes	No	No	Novel exon	g4159.t1
163	CHLCTEEK	C2H2 Transcription factor XP_001243942.1	Yes	No	No	Novel exon	g1846.t1
164	DLYGNIVMSGGTTMYPGIADR		Yes	No	No	Novel exon	g6034.t1
165	DWVNACSSAFDIPVIEDFNK		Yes	No	No	Novel exon	g4402.t1
166	EGNPLPTLAKVQPGNLSWR		Yes	No	No	Novel exon	g3881.t1
167	ELKVCIAQDNGTTFER		Yes	No	No	Novel exon	g5720.t1
168	ELSSIVCKYTVVDSMRPDVAELVK		Yes	No	No	Novel exon	g3548.t1
169	FCGAACHDGGNNHSDTC		Yes	No	No	Novel exon	g2670.t1
170	FHQGPVEPQFGPGR		Yes	No	No	Novel exon	g4095.t1
171	FIDSNFSFEVVEEFFQKFLK	Overlaps low-complexity region	Yes	No	No	Novel exon	g2650.t1
172	FYSQRNMYLCGFTLFLSLILNR		Yes	No	No	Novel exon	g1941.t1
173	GDNNRAWKRRRR		Yes	No	No	Novel exon	g4410.t1
174	GLLVPTTTTDQELQHLR		Yes	No	No	Novel exon	g625.t1
175	GPSGVGESNPPTPASGATPTPSGGDKPMES	Overlaps low-complexity region	Yes	No	No	Novel exon	g1291.t1
176	KLSFPAMSGFDQATSLGAWK		Yes	No	No	Novel exon	g6924.t1
177	LASTPQSADSLR		Yes	No	No	Novel exon	g4042.t1
178	LFASEAHSGTPFFNSIFMFEGYSVR		Yes	No	No	Novel exon	g4077.t1
179	LLESDASGQNENVAEK		Yes	No	No	Novel exon	g734.t1
180	LTGHMGPQYSPAASYR		Yes	No	No	Novel exon	g113.t1
181	MREGWLDIDAMR		Yes	No	No	Novel exon	g1763.t1
182	NMYLCGFTLFLSLILNR		Yes	No	No	Novel exon	g1941.t1
183	PDALVPFAPNAK		Yes	No	No	Novel exon	g2008.t1
184	PSDTAAVVVHEFEK		Yes	No	No	Novel exon	g4410.t1
185	QDDESGITPVCTFPIMEPEKPEK	Crosses splice boundary	Yes	No	No	Novel exon	g827.t1
186	QMSVYGHRCLGPRGR		Yes	No	No	Novel exon	g5348.t1
187	QTTPGSHGLAR	C2H2 Transcription factor XP_001243942.1	Yes	No	No	Novel exon	g1846.t1
188	SIGGGDDIEALDISRELAPK		Yes	No	No	Novel exon	g882.t1
189	SIMHTPSTRPSK		Yes	No	No	Novel exon	g3905.t1
190	SVAELVIGEIIALAR		Yes	No	No	Novel exon	g523.t1
191	SVPGMEHTKQLADVLYGR		Yes	No	No	Novel exon	g466.t1
192	TFCRLTALASWRNEYILR		Yes	No	No	Novel exon	g6829.t1
193	TWDADSPGPISIQLLR		Yes	No	No	Novel exon	g4530.t1
194	VAALVFRPNEVDTV		Yes	No	No	Novel exon	g3195.t1
195	VFMFEFPERPGALAR		Yes	No	No	Novel exon	g3877.t1
196	VIILEATEIAGGASGK		Yes	No	No	Novel exon	g3249.t1
197	VIQSSDVFLYGGGLYSFFE		Yes	No	No	Novel exon	g2008.t1
198	VNAAFETNCKGENPDEPTM		Yes	No	No	Novel exon	g2903.t1
199	WLHECIFEKYWTKPSR		Yes	No	No	Novel exon	g3881.t1
200	YNTRFIALKFAYLGQR		Yes	No	No	Novel exon	g1763.t1

	Peptide	Notes	AUG	Six	EST	Peptide Classification	Novel Gene Model/ locati
201	YRQTMMYTATMPSAVERIARK		Yes	No	No	Novel exon	g5773.t1
202	AELEFEMITDMVNR		Yes	No	Yes	Novel exon	g5873.t1
203	AGEEVSVTGLK		Yes	No	Yes	Novel exon	g3529.t1
204	DANANPIEGSSQPSDK		Yes	No	Yes	Novel exon	g2221.t1
205	GAPFAGATNR		Yes	No	Yes	Novel exon	g2248.t1
206	GESVCLDR		Yes	No	Yes	Novel exon	g5873.t1
207	GGHNVLVDNFLNAQYFSEISIGNPPQNFK		Yes	No	Yes	Novel exon	g1628.t1
208	GLEFTEFIPDGEWQAK		Yes	No	Yes	Novel exon	g3529.t1
209	KYDSSASSTYK		Yes	No	Yes	Novel exon	g1628.t1
210	LFYGVWGDTNGGTLVGETSVSLAQACFPK		Yes	No	Yes	Novel exon	g6812.t1
211	RRPTMDEEEETDLMFVRK		Yes	No	Yes	Novel exon	g1216.t1
212	TGGHNVLVDNFLNAQYFSEISIGNPPQNFK		Yes	No	Yes	Novel exon	g1628.t1
213	TLSDLADLFLDEIK		Yes	No	Yes	Novel exon	g1216.t1
214	VAVIGGSIDYTGAPYFSAMASAK		Yes	No	Yes	Novel exon	g62.t1
215	YFSEISIGNPPQNFK		Yes	No	Yes	Novel exon	g1628.t1
216	YPSLIDATAEELQDGLSR		Yes	No	Yes	Novel exon	g3107.t1
217	LSGQEDDVDIQEAIR		Yes	No	No	Peptides out of frame of existing annotation	g6772.t1
218	TDTRHYLYWVNSTLQSANR		Yes	No	No	Peptides out of frame of existing annotation	g2739.t1
219	CFTSVDLVNAYSMR		Yes	No	Yes	Peptides out of frame of existing annotation	g3107.t1
220	FIIFSNTDTRHYLYWVNSTLQSANR		Yes	No	Yes	Peptides out of frame of existing annotation	g2739.t1
221	HYLYWVNSTLQSANR		Yes	No	Yes	Peptides out of frame of existing annotation	g2739.t1
222	IAGMAAAYGNGDYR		Yes	No	Yes	Peptides out of frame of existing annotation	g3496.t1
223	LYWVNSTLQSANR		Yes	No	Yes	Peptides out of frame of existing annotation	g2739.t1
224	QDITSMPQCAQSPVLEAISSSK		Yes	No	Yes	Peptides out of frame of existing annotation	g6526.t1
225	WVNSTLQSANR		Yes	No	Yes	Peptides out of frame of existing annotation	g2739.t1
226	SDFALGADR		No	No	Yes	Peptides out of frame of existing annotation	CPSG_01623
227	YWVNSTLQSANR		Yes	No	Yes	Peptides out of frame of existing annotation	g2739.t1
228	**GASQQDNEGGYPR**		Yes	Yes	No	Peptides part of putative transposons	tr1
229	**QSQWDKDQNEK**		Yes	Yes	No	Peptides part of putative transposons	tr1
230	**TPVKPLPK**		Yes	Yes	No	Peptides part of putative transposons	tr1
231	FVTAGFDAR		No	No	Yes	Peptide too short for confident description	Unknown
232	AIETLNTGR		Yes	No	No	Peptide too short for confident description	g1216.t1
233	AKYESDPQK		Yes	No	No	Peptide too short for confident description	g6903.t1
234	ARGEGDDSLNR		Yes	No	No	Peptide too short for confident description	g4460.t1
235	AVIFEFLEAIR		Yes	No	No	Peptide too short for confident description	g1354.t1
236	CHPPVER		Yes	No	No	Peptide too short for confident description	g4641.t1
237	CLGGCDPK		Yes	No	No	Peptide too short for confident description	g4506.t1
238	DFLDEASK		Yes	No	No	Peptide too short for confident description	g2126.t1
239	DVNTLLNR		Yes	No	No	Peptide too short for confident description	g2739.t1
240	EDLLPVFSFK		Yes	No	No	Peptide too short for confident description	g3877.t1
241	EEGQNVTIPGK		Yes	No	No	Peptide too short for confident description	g3883.t1
242	EGGEGGVGVGSAR		Yes	No	No	Peptide too short for confident description	g6058.t1
243	ELAESVIAK		Yes	No	No	Peptide too short for confident description	g6772.t1
244	ELFTAIAK		Yes	No	No	Peptide too short for confident description	g3883.t1
245	EPIVIPR		Yes	No	No	Peptide too short for confident description	g6613.t1
246	FFANEQR		Yes	No	No	Peptide too short for confident description	g1643.t1
247	FKDGLDAGFK		Yes	No	No	Peptide too short for confident description	g2693.t1
248	FLNDILK		Yes	No	No	Peptide too short for confident description	g4263.t1
249	FLSTSAK		Yes	No	No	Peptide too short for confident description	g3322.t1
250	FSVDLPR		Yes	No	No	Peptide too short for confident description	g3146.t1
251	GFPQSPENR		Yes	No	No	Peptide too short for confident description	g4132.t1
252	GIDFFASR		Yes	No	No	Peptide too short for confident description	g4178.t1
253	GISPSEVCSR		Yes	No	No	Peptide too short for confident description	g6099.t1
254	GQTPSLKF		Yes	No	No	Peptide too short for confident description	g3907.t1
255	HEFRQIFFEM		Yes	No	No	Peptide too short for confident description	g4461.t1
256	HPGYLTAER		Yes	No	No	Peptide too short for confident description	g366.t1
257	IAPELYLK		Yes	No	No	Peptide too short for confident description	g1354.t1
258	IFRGMLDLLR		Yes	No	No	Peptide too short for confident description	g4159.t1
259	KITPCVEK		Yes	No	No	Peptide too short for confident description	g2934.t1
260	LFDILDER		Yes	No	No	Peptide too short for confident description	g4428.t1
261	LGSPAVR		Yes	No	No	Peptide too short for confident description	g2121.t1
262	LIELLLER		Yes	No	No	Peptide too short for confident description	g967.t1
263	LLQALNESR		Yes	No	No	Peptide too short for confident description	g2650.t1
264	LQVGQPQTR		Yes	No	No	Peptide too short for confident description	g873.t1
265	LSGLFSK		Yes	No	No	Peptide too short for confident description	g2356.t1
266	LSSMSLVNR		Yes	No	No	Peptide too short for confident description	g6829.t1
267	MAALKPAFVR		Yes	No	No	Peptide too short for confident description	g6092.t1
268	MAPPDVR		Yes	No	No	Peptide too short for confident description	g2485.t1
269	NIVISHK		Yes	No	No	Peptide too short for confident description	g967.t1
270	PKWDDELYER		Yes	No	No	Peptide too short for confident description	g6973.t1
271	PQLEPYFK		Yes	No	No	Peptide too short for confident description	g4790.t1
272	QAIPLMR		Yes	No	No	Peptide too short for confident description	g417.t1
273	QIVNVPSFIVR		Yes	No	No	Peptide too short for confident description	g3620.t1
274	RHKMEIAHR		Yes	No	No	Peptide too short for confident description	g4806.t1
275	SDGNEIAMKR		Yes	No	No	Peptide too short for confident description	g3940.t1
276	SIPHETR		Yes	No	No	Peptide too short for confident description	g510.t1
277	TELGQLR		Yes	No	No	Peptide too short for confident description	g947.t1
278	TETLLEEGR		Yes	No	No	Peptide too short for confident description	g3883.t1
279	TIIAAWR		Yes	No	No	Peptide too short for confident description	g5915.t1
280	TINLPAR		Yes	No	No	Peptide too short for confident description	g471.t1
281	TKPIWTR		Yes	No	No	Peptide too short for confident description	g5616.t1
282	TYKLPGFEK		Yes	No	No	Peptide too short for confident description	g2073.t1
283	VQNTLPDLR		Yes	No	No	Peptide too short for confident description	g4806.t1
284	VVSINGGR		Yes	No	No	Peptide too short for confident description	g2407.t1
285	WRRLVELGVLK		Yes	No	No	Peptide too short for confident description	g2126.t1
286	YALEGLR		Yes	No	No	Peptide too short for confident description	g2093.t1
287	YIDYTKTK		Yes	No	No	Peptide too short for confident description	g4951.t1
288	YMPAVGDIK		Yes	No	No	Peptide too short for confident description	g3474.t1

APPENDIX C

COVERAGE OF IDENTIFIED PROTEINS
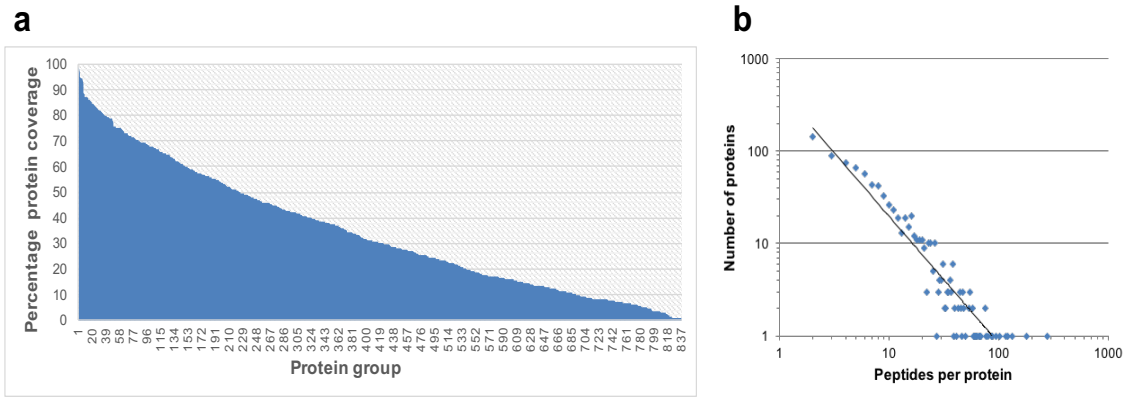
Coverage of identified proteins. a) Percentage of protein coverage by protein group. The average protein coverage in this study was 35.1%, with a range of 99.3% to 0.7% coverage.

b) A log-log plot of the number of peptides per protein. Each point can represent many proteins with the same number of peptides. There was average of 12.4 peptides per protein with a minimum of 2 peptides and maximum of 278 peptides per protein. The distribution of peptides per protein is linear, as seen by the best fit line.

APPENDIX D

MEAN CAZYME NORMALIZED SPECTRAL ABUNDANCE FACTOR (NSAF)

VALUES FROM ALL *IN VITRO* AND *IN VIVO* SAMPLES.

175

Protein name	Uniprot Accession Number	CAZyme Families	In vitro Coccidioides (Mycelial Form)						In vitro Coccidioides (Spherule Form)					
			Converse + Tamol	Converse + Survanta	RPMI+ FBS	RPMI+Survanta	RPMI alone	Converse alone	Converse + Tamol	Converse + Survanta	RPMI+ FBS	RPMI+Survanta	RPMI alone	Converse alone
Obg-like ATPase 1	A0A0E1RUI9_COCIM	GH72	5.06E-05	0.00064613	0.0001315	5.77E-05	0.00039087	0.0006195	0.0004686	0.000686723	0.000369	0.001579493	0.0007699	0.00055065
FAD binding domain-containing protein	A0A0J6ETK6_COCPO	AA7	0	0	0	0	0	1.435E-05	0	0	0	0	0	0
Carboxylic ester hydrolase (EC 3.1.1.-)	A0A0J6EUR6_COCPO	CE10	0.00082	0.0001285	0	0	0.00022983	3.496E-05	0	0	0	0	0	0
Isoamyl alcohol oxidase	A0A0J6F564_COCPO	AA7	3.52E-05	0	0.0002106	0	0	4.751E-05	0	0	0	0	0	0
STT3 protein	A0A0J6F5R9_COCPO	GT66	2.84E-05	7.8707E-06	9.227E-06	6.47E-05	0	4.036E-05	0.000115	0.00008691	0	8.61867E-05	0	0.00004104
Laccase	A0A0J6F875_COCPO	AA1	0	0	0	0	0	0	0.000871	0.000369347	0.000684	0.000913887	0.0007073	0.00043417
Uncharacterized protein	A0A0J6FBL6_COCPO	GH18	0	8.4357E-06	0	0	0	0.0001024	0	2.66133E-05	0	4.61047E-05	1.978E-05	2.1992E-05
Initiation-specific alpha-1,6-mannosyltran	A0A0J6FEW7_COCPO	GT32	0	0	3.629E-05	0	0	0	0	0	0	0.00008284	0	0
Uncharacterized protein	A0A0J6FGV4_COCPO	GH17	0.001239	0.00026282	0	0.002968	0.00035465	0.0001275	0	0	0	0	0	0
N-acetylglucosamine-6-phosphate deacet	A0A0J6FHV0_COCPO	CE9	0	0.00038405	4.478E-05	0.000105	0	0.0001716	0	0	0	0	1.493E-05	0
Glutamyl-tRNA synthetase	A0A0J6FIX1_COCPO	GT22	2.82E-05	0.00056983	0.0001741	9.64E-05	0.00039508	0.0003637	0.0002612	0.000633537	0.000402	0.001398267	0.0007325	0.00052292
Cytoplasm protein	A0A0J6FJD1_COCPO	GH5	0.000228	8.7635E-05	1.696E-05	0.000691	4.9587E-05	3.305E-05	0.0001056	1.14083E-05	0	4.94967E-05	0.0000424	1.8855E-05
Trehalose-phosphatase	A0A0J6FLG5_COCPO	GT20	2.19E-05	7.6304E-05	1.378E-05	0	0	3.657E-05	2.146E-05	6.49067E-05	0	0.000181557	0.0001103	7.8278E-05
Glycogen debranching enzyme	A0A0J6FLH7_COCPO	GH13	8.3E-05	0.00002558	4.284E-05	1.5E-05	8.5517E-05	0.0001462	4.003E-05	0.00005188	0	0.000069472	3.856E-05	8.5743E-05
Glycogen debranching enzyme	A0A0J6FLH7_COCPO	GH133	8.3E-05	0.00002558	4.284E-05	1.5E-05	8.5517E-05	0.0001462	4.003E-05	0.00005188	0	0.000069472	3.856E-05	8.5743E-05
NADP-dependent leukotriene B4 12-hydro	A0A0J6FPY7_COCPO	GH36	0.000189	0.00198547	0.0003164	6.53E-05	0.00408973	0.0013754	0.0027463	0.003175733	0.002585	0.003154767	0.003338	0.00209528
Protein HIR	A0A0J6FR55_COCPO	GT15	0	0	2.442E-05	0	0	0	0	0	0	0	0	0
Chitinase 1, variant	A0A0J6I9W4_COCPO	GH18	0.002943	0.00094588	0.0009931	0.001565	0	0.0008803	0	0	0	0	0	0
1,3-beta-glucanosyltransferase (EC 2.4.1.-	A0A0J6XX95_COCIT	GH72	8.92E-05	0.00075153	0.0001884	0.000356	0.00004068	0.0006525	4.513E-05	1.94987E-05	0	0.000285989	0.0001701	0.0001289
Cell wall synthesis protein	A0A0J6YCH2_COCIT	GH132	6.04E-05	2.588E-05	1.517E-05	0	0.00042818	6.963E-05	0	0	0	1.77117E-05	0	0
Gram-negative bacteria-binding protein 1	A0A0J6YEV0_COCIT	AA5	0	0	9.436E-06	0	0	0	0	0	0	0	0	0
Gram-negative bacteria-binding protein 1	A0A0J6YEV0_COCIT	GH16	0	0	9.436E-06	0	0	0	0	0	0	0	0	0
Mannan polymerase II complex ANP1 sub	A0A0J6YN81_COCIT	GT62	0	1.3444E-05	3.152E-05	0	0	5.425E-05	0	0	0	0.000018271	0	0
Ecm33	A0A0J7BE11_COCIT	GH0	0.002006	0.00136297	0.0021108	0.003174	0.00191927	0.0020542	0.0019692	0.000828387	0	0.001150307	0.0014186	0.00133991
Mannan endo-1,6-alpha-mannosidase (EC	A0A0J8QVM6_COCIT	GH76	0	0	9.879E-05	0.000149	0.00015293	4.858E-05	0	0	0	0.000048071	0	0
alpha-1,2-Mannosidase (EC 3.2.1.-)	A0A0J8RCD3_COCIT	GH47	0.000167	1.1897E-05	0	0	0	5.134E-05	0	0	0	0	0	0
Cystathionine gamma-synthase (O-succin	A0A0J8S1Z0_COCIT	CE4	0	0.00017172	0.0000481	0	0.00026125	0.0001181	0.0002497	0.000151017	0	0.000224507	8.018E-05	0.00007131
Cell wall alpha-1,3-glucan synthase ags1	A0A0J8UAT3_COCIT	GH13	0	5.633E-05	2.448E-05	8.59E-06	0	1.797E-05	3.05E-05	0.000105167	8.12E-05	0.000162547	4.801E-05	4.7685E-05
Cell wall alpha-1,3-glucan synthase ags1	A0A0J8UAT3_COCIT	GT5	0	5.633E-05	2.448E-05	8.59E-06	0	1.797E-05	3.05E-05	0.000105167	8.12E-05	0.000162547	4.801E-05	4.7685E-05
Dolichol-phosphate mannose synthase	A0A0J8UCK8_COCIT	GT2	0.000407	0.00020353	0.0005246	8.2E-05	6.5647E-05	0.0004148	0	0	0	5.44123E-05	4.678E-05	5.2003E-05
Glucosidase I, putative (EC 3.2.1.106)	C5NZS8_COCP7	GH63	9.91E-05	0.00029404	0.0001425	8.34E-05	0	0.0003458	0	0.00004262	0	0.000120193	3.168E-05	2.9075E-05
Glycosyl hydrolase family 16 protein	C5P0X8_COCP7	GH16	0.000607	1.1162E-05	0	0.000635	0	0.0004098	0	0	0	0	0	0
Flavodoxin domain containing protein	C5P137_COCP7	AA6	0.003594	0.0030266	0.0026284	0.002667	0.00572273	0.0033363	0.0026748	0.003485367	0.001628	0.0020226	0.0028816	0.00316003
S-formylglutathione hydrolase (EC 3.1.2.1	C5P172_COCP7	CE1	0.00049	0.00117233	0.0002174	0	0.00244913	0.0009311	0.0021136	0.001337733	0.002355	0.001920567	0.0028577	0.00268803
Uncharacterized protein	C5P1K3_COCP7	GH76	0	2.7668E-05	0	0	0	0	0	0	0	0	0	0
Dolichyl-phosphate-mannose--protein ma	C5P2G5_COCP7	GT39	2.67E-05	7.3967E-06	2.601E-05	0.000122	0	5.37E-05	2.701E-05	1.16677E-05	0	6.98917E-05	1.735E-05	0
Valyl-tRNA synthetase, mitochondrial, pu	C5P2N7_COCP7	GT13	0	0.00010741	0.0001054	0	0.00013725	0.0001626	7.722E-05	8.34067E-05	7.29E-05	0.000100569	0.000124	8.2707E-05
D-lactate dehydrogenase, putative	C5P4E8_COCP7	AA4	3.63E-05	0.00032435	4.711E-05	0	0.000135	0.0001872	0.0002624	0.00034564	6.84E-05	0.000583593	0.0003338	0.00026982
Glycogen [starch] synthase (EC 2.4.1.11)	C5P4H1_COCP7	GT3	2.79E-05	0.00016996	7.258E-05	9.55E-05	0	0.0003116	0	5.02427E-05	0	0.000155813	0	0.00006662
Uncharacterized protein	C5P4K7_COCP7	AA11	0	0	0	0	0.00019193	0	0.0002548	0.000234987	0	0.0002039	0.0028206	0.00176745
Glucosidase II alpha subunit, putative (EC	C5P4Q6_COCP7	GH31	0.000158	0.00015133	6.733E-05	0	9.5027E-05	0.0001252	0.0002726	0.000135903	8.01E-05	0.000156529	0.0001473	3.9697E-05
Aminopeptidase I zinc metalloprotease fa	C5P4W6_COCP7	GH71	0.003837	0.00173073	0.0013197	0.000516	0.00126622	0.0015111	0.0011937	0.0008925	0.002461	0.000961973	0.0012172	0.00100582
CNH domain containing protein	C5P579_COCP7	GT31	0	3.1944E-05	5.34E-06	0	0	1.711E-05	0	0	0	0	0	0
Heat shock protein 78, mitochondrial, put	C5P5L4_COCP7	GH6	0	0.00061058	0.0003457	0	0.00035955	0.0003612	0.001612	0.004230133	0.000712	0.001902533	0.0019468	0.00192703
1,3-beta-glucanosyltransferase (EC 2.4.1.-	C5P695_COCP7	GH72	0.001604	0.00024134	0.0003013	0.003186	0.00011854	0.0003879	0	0	0	0	0	0
Lipase, putative	C5P6V3_COCP7	CE10	0.000102	0.00059154	0.0002325	0.001179	0.00051415	0.0004169	0.0001552	0.00013409	0.000393	0.001050503	0.0001661	0.00007387

176

			In vitro Coccidioides (Mycelial Form)						In vitro Coccidioides (Spherule Form)					
Protein name	Uniprot Accession Number	CAZyme Families	Converse + Tamol	Converse + Survanta	RPMI+ FBS	RPMI+Survanta	RPMI alone	Converse alone	Converse + Tamol	Converse + Survanta	RPMI+ FBS	RPMI+Survanta	RPMI alone	Converse alone
Alpha-1,4 glucan phosphorylase (EC 2.4.1.	C5P6Y0_COCP7	GT35	0.000648	0.00069263	0.0005738	0.000206	0.00097928	0.0012574	0.0006183	0.000746987	0.001566	0.00122619	0.000712	0.00074841
Uncharacterized protein	C5P7R8_COCP7	AA7	0.003393	0.00034522	0.0097361	0.004908	0.00015442	0.00105	0	0	0	0.000074633	0.0001065	0
Glycosyl transferase, group 1 family prote	C5P7T1_COCP7	GT4	0.000883	0.0022918	0.0014499	0.000802	0.00132363	0.0026877	0.0009569	0.001102117	0.001199	0.00149695	0.0007288	0.00066023
Beta-1,3-glucanase, putative (EC 3.2.1.58	C5P839_COCP7	GH55	0.005608	0.00142433	0.0051244	0.004008	0.00053809	0.0011899	0.0002568	0.000460373	0	0.000117126	0.0004827	0.00034416
Uncharacterized protein	C5P9X5_COCP7	GH75	0	0.00010232	2.185E-05	0	0	0.0000376	0	0	0	0	0	0
Heat shock protein HSP98, putative	C5PAL0_COCP7	GH6	0.000152	0.00097147	0.0006416	4.94E-05	0.0011369	0.0010806	0.0033075	0.003838467	0.002054	0.0039815	0.0014649	0.00203288
WSC domain containing protein	C5PAT9_COCP7	AA2	2.72E-05	0	0	6E-05	0	0	0	0	0	0	0	0
DNA-directed RNA polymerase subunit be	C5PAW2_COCP7	GT0	0	9.2724E-05	7.209E-05	1.81E-05	2.8909E-05	0.0001754	4.811E-05	4.93027E-05	5.52E-05	8.25917E-05	0	2.2901E-05
D-lactate dehydrogenase, putative	C5PB22_COCP7	AA7	0	0.00020413	9.62E-05	3.75E-05	8.1643E-05	0.0001326	0.0008475	0.000631783	0	0.0013595	0.0002523	0.00038289
Acetyltransferase, GNAT family protein	C5PBE6_COCP7	GH3	2.35E-05	4.6267E-05	0	0	0	0.0001348	0	0	0	0	0	0
Endochitinase 1, putative (EC 3.2.1.14)	C5PBT0_COCP7	GH18	0.000159	0.00028223	0.0001677	0.001278	0.00053724	0.000211	4.019E-05	0.000190987	0	0.000405412	9.034E-05	8.6083E-05
FAD binding domain containing protein	C5PCE3_COCP7	AA7	0.004476	0.00177007	0.0086621	0.007078	0.00124489	0.0021599	0	0	0	0.000502633	1.26E-05	0
Class III chitin synthase	C5PCI6_COCP7	GT2	0	6.1187E-06	0	0.000126	0	1.662E-05	0	0	0	0	0	0
Actin-related protein 3, putative	C5PD17_COCP7	GT22	4.55E-05	0.00092688	0.0003401	5.19E-05	0.00087523	0.0006748	4.606E-05	0.00064274	8.79E-05	0.000373803	0.0004487	0.00029445
26S proteasome regulatory subunit RPN1	C5PD33_COCP7	GT0	0.000135	5.6632E-05	9.346E-05	5.04E-05	0.00012896	0.0002733	4.478E-05	8.70633E-05	0.000162	0.000183675	7.19E-05	7.9937E-05
alpha-1,2-Mannosidase (EC 3.2.1.-)	C5PEE6_COCP7	GH47	0	0	2.718E-05	0	0	0	0	0.00001219	0	0	0	0
ATP-dependent RNA helicase, putative	C5PEF7_COCP7	GH37	7.95E-05	0.00046092	0.0001905	0.000178	0.0002149	0.0006029	7.911E-05	0.0001709	0	0.000221108	0.0002016	0.00044511
Lipase, putative (EC 3.1.1.3)	C5PF68_COCP7	CE10	0.00076	0.0001097	0.0010722	0.000534	0.00010371	0.0002149	2.818E-05	0.000012173	0	0	0	0
1,3-beta-glucanosyltransferase (EC 2.4.1.	C5PFR4_COCP7	CBM43	0.002253	0.00052318	0.0003312	0.001983	0.00037496	0.0006737	0	0.00006603	0	0.000157358	2.454E-05	8.1847E-05
1,3-beta-glucanosyltransferase (EC 2.4.1.	C5PFR4_COCP7	GH72	0.002253	0.00052318	0.0003312	0.001983	0.00037496	0.0006737	0	0.00006603	0	0.000157358	2.454E-05	8.1847E-05
Trehalose-6-phosphate synthase (EC 2.4.1	C5PFT3_COCP7	GT20	7.85E-05	4.3507E-05	2.55E-05	4.47E-05	3.08E-05	0.0001201	0.0001191	0.000190717	0	0.000192099	0.0002628	0.00032355
6-hydroxy-D-nicotine oxidase, putative (E	C5PGK3_COCP7	AA7	0.000611	0.00110784	0.0007568	0.00043	0.00098568	0.0011207	8.477E-05	0.00022609	0.000385	0.000609683	0.0001225	0.00018157
6-phosphogluconate dehydrogenase, deca	C5PH56_COCP7	GT30	0.001395	0.00249263	0.002464	0.001176	0.00530267	0.0036073	0.0019499	0.00161407	0.003616	0.0025732	0.0015449	0.00176914
Uncharacterized protein	C5PH87_COCP7	GH128	0	0	0	0	0.0004338	0	0	0.00003418	0.000302	5.93167E-05	7.622E-05	0.00011298
Glycolipid 2-alpha-mannosyltransferase f	C5PI37_COCP7	GT15	0.000421	0.00011728	0.0003733	0.000558	0.00020006	0.0001055	0.0001903	0.000123337	0	0.00030016	6.112E-05	0.00013589
Thioredoxin reductase (EC 1.8.1.9)	C5PIA5_COCP7	GT4	0.001164	0.00039557	0.0002153	0.000315	0.00065337	0.0008055	0.0021444	0.004599767	0.00082	0.001130163	0.0025067	0.00154187
Cutinase family protein	C5PII3_COCP7	CE5	8.94E-05	2.4777E-05	0	0	0	0	0	0	0	0	0	0
Peroxidase (EC 1.11.1.-)	C5PJM7_COCP7	AA2	0.000797	0.00224787	0.0009551	0.000975	0.00173873	0.0014824	0.0025622	0.0037796	0.006621	0.004358433	0.0036844	0.00206223
Endochitinase 1 (EC 3.2.1.14) (Compleme	CHI1_COCIM	GH18	0.052296	0.00581027	0.0333725	0.01929	0.0041533	0.0127023	0.0013405	0.000236355	0	7.03467E-05	0.0015507	0
1,4-alpha-glucan-branching enzyme	E9CT92_COCPS	CBM48	0.000155	0.00039974	3.777E-05	3.31E-05	3.2418E-05	0.0004051	0	0	0	0	2.833E-05	2.0998E-05
1,4-alpha-glucan-branching enzyme	E9CT92_COCPS	GH13	0.000155	0.00039974	3.777E-05	3.31E-05	3.2418E-05	0.0004051	0	0	0	0	2.833E-05	2.0998E-05
Uncharacterized protein	E9CUG2_COCPS	AA7	0.000497	9.531E-05	0.0005206	0.001221	0.00062012	9.305E-05	0	0	0	0	0	0
Beta-glucosidase	E9CUL1_COCPS	GH17	0.002871	0.00098499	0.0004234	0.001874	0.00092754	0.0010222	0.0028036	0.002715483	0.002999	0.002064033	0.0033064	0.00224997
Glucoamylase (EC 3.2.1.3) (1,4-alpha-D-g	E9CVY5_COCPS	CBM20	0.002036	0.00040911	0.0019874	0.00104	0	0.0004746	0	0	0	0.000141815	0	0
Glucoamylase (EC 3.2.1.3) (1,4-alpha-D-g	E9CVY5_COCPS	GH15	0.002036	0.00040911	0.0019874	0.00104	0	0.0004746	0	0	0	0.000141815	0	0
Uncharacterized protein	E9CWH3_COCPS	CE10	0	3.6636E-05	0	2.34E-05	0	0	0	1.79523E-05	0	0.00012973	3.336E-05	1.4835E-05
Beta-glucosidase (EC 3.2.1.21)	E9CWI7_COCPS	GH3	0.000329	0.00070271	0.0001114	0	0	0.0008018	0	0	0	0	0	0
Uncharacterized protein	E9CXC8_COCPS	GH16	0.007502	0.00126449	0.0007071	0.006779	0.0011417	0.0016231	0.0003552	0.002725407	0	0.0003688	0.0012614	0.00035503
Adenylosuccinate synthetase (AMPSase)	E9CXU9_COCPS	GT47	0.000146	0.00039415	0.0002319	5.42E-05	0.00090904	0.0004446	0.0005925	0.000536907	9.19E-05	0.001329597	0.000199	0.00030941
Beta-glucosidase (EC 3.2.1.21)	E9CZV7_COCPS	GH3	0.007889	0.00205717	0.0009900	0.005857	0.00053867	0.0018614	0.0006456	0.000924553	0	0.000502497	9.061E-05	0.00020147
DUF1237 domain-containing protein	E9D1F3_COCPS	GH125	0.001809	0.00048532	0.0017564	0.001308	0.00024156	0.0004552	0.0002934	0.00075911	0	0.000325933	3.534E-05	0.00018333
Dolichyl-phosphate-mannose-protein mar	E9D4S5_COCPS	GT39	4.32E-05	5.8907E-06	0	9.69E-05	0	7.924E-06	6.452E-05	5.57533E-05	0	4.83767E-05	2.072E-05	0
Raffinose synthase Sip1	E9D6H0_COCPS	GH36	7.21E-05	2.426E-05	7.11E-06	0	0	8.176E-05	0	0	0	0	0	0
Choline oxidase	E9D8U5_COCPS	AA3	0	0.00030013	7.144E-05	4.18E-05	0	4.641E-05	0.0016343	0.006516167	0.003261	0.0025408	0.0030233	0.00165584
Glycosidase (EC 3.2.-.-)	E9D9A5_COCPS	CBM18	0.000607	0.00019624	0.0001035	0.000208	0.00012455	0.0002552	4.606E-05	0	0	1.72667E-05	0	3.2888E-05
Glycosidase (EC 3.2.-.-)	E9D9A5_COCPS	GH16	0.000607	0.00019624	0.0001035	0.000208	0.00012455	0.0002552	4.606E-05	0	0	1.72667E-05	0	3.2888E-05
UDP-glucose:glycoprotein glucosyltransfe	E9D9R1_COCPS	GT24	0	0.00021599	1.741E-05	1.53E-05	6.9277E-05	7.39E-05	0	5.85733E-06	0	0.000090085	7.401E-05	0
Trehalase (EC 3.2.1.28) (Alpha-trehalose	E9DA57_COCPS	GH37	0	0.00015266	1.772E-05	0	2.4875E-05	3.876E-05	0	0	0	4.13833E-05	3.545E-05	0.00003941
Uncharacterized protein	E9DBB7_COCPS	GH18	0	0.00077303	0.0001761	0	0.00093542	0.000517	0.0002438	0.001102727	0	0.00136544	0.0007258	0.00055323
Acid trehalase	E9DEJ8_COCPS	GH65	0.002408	0.00048	0.0002128	0.000314	4.3747E-05	0.0005541	0.0006724	0.001155413	0.0002	0.00061627	4.83E-05	0.0000537
Alpha-mannosidase (EC 3.2.1.-)	E9DGJ4_COCPS	GH38	0.003806	0.0025116	0.0055051	0.00458	0.00099758	0.0024144	0.0003347	0.00008033	0.000204	0.000118497	0.0002565	0.00026233
Glucose-6-phosphate 1-dehydrogenase (E	E9DHZ4_COCPS	CBM13	3.91E-05	0.00096595	0.0003175	4.46E-05	0.00048136	0.0008898	7.911E-05	0.00012161	7.38E-05	0.000437917	0.0001016	0.00016947
Uncharacterized protein	E9DJZ5_COCPS	AA7	0	0	0	0	4.5017E-05	0	0.0001225	1.76433E-05	0	0.000332693	6.557E-05	0.00005832
Beta-hexosaminidase 1 (EC 3.2.1.52) (Bet	HEX1_COCPS	GH20	0.006312	0.00491877	0.0063526	0.006159	0.00054324	0.0045418	0.0004069	0.00131837	0	0.000381303	1.089E-05	0.00016947
ATP synthase subunit alpha	J3KD83_COCIM	GT4	0.001983	0.00589937	0.0048458	0.002252	0.00826073	0.0052388	0.0054457	0.007231533	0.007344	0.005860333	0.0082658	0.00884963
Endo-1,3-beta-glucanase	J3KEN5_COCIM	GH81	0.000169	0.00010768	0.0002043	0	0	8.82E-05	0	0	0	0	0	0
Trehalose-phosphatase	J3KKI8_COCIM	GT20	0	1.1908E-05	0	2.45E-05	0	1.602E-05	8.696E-05	0.00008453	4.15E-05	7.25013E-05	3.491E-05	0
Arp2/3 complex subunit Arc16	J3KLS0_COCIM	CE4	0	0.0001044	2.076E-05	0	0.00011656	0	0	0	0	0.000261783	0	0
Mannosyl-oligosaccharide alpha-1,2-man	MNS1B_COCPS	GH47	0.01332	0.00256953	0.0121054	0.011598	0.0018618	0.0044029	0.0006219	0.003853603	0.000777	0.00201968	0.0004818	0.00166753
Endo-1,3-beta-glucanase (Fragment)	Q0MYQ6_COCPO	GH81	0.010983	0.00376437	0.0139245	0.005038	0.00041182	0.0034535	0	2.42107E-05	0.000374	0	8.759E-05	0
Class V chitin synthase	Q71CG1_COCPO	GT2	0	6.2363E-06	0	0	0	3.324E-06	0	4.91867E-06	0	0	0	0

APPENDIX E

TOP 150 HUMAN GLYCAN M/Z

	mz	p	q	Cocci Lysate	True pos	False Pos	False Neg	True Neg
	NA	**NA**	**NA**					
1	726.274	9.68E-07	0.7090899	0	11	2	9	57
2	393.05	4.35E-06	1	0	8	0	12	59
3	438.199	2.10E-05	1	0	10	3	10	56
4	803.272	2.44E-05	1	1	7	0	13	59
5	404.231	6.37E-05	1	4	10	3	10	56
6	1240.7419	0.000101	1	0	9	3	11	56
7	577.773	0.00012898	1	0	6	0	14	59
8	908.386	0.00012898	1	0	6	0	14	59
9	612.951	0.00012898	1	0	6	0	14	59
10	1077.729	0.00012898	1	0	6	0	14	59
11	650.755	0.00012898	1	0	6	0	14	59
12	614.026	0.00012898	1	0	6	0	14	59
13	1157.989	0.00012898	1	0	6	0	14	59
14	723.267	0.00012898	1	0	6	0	14	59
15	458.531	0.00012898	1	0	6	0	14	59
16	805.011	0.00012898	1	0	6	0	14	59
17	497.176	0.00012898	1	0	6	0	14	59
18	857.634	0.00012898	1	0	6	0	14	59
19	467.495	0.00012898	1	0	6	0	14	59
20	588.126	0.00012898	1	0	6	0	14	59
21	1216.0861	0.00012898	1	0	6	0	14	59
22	884.503	0.00014165	1	0	8	2	12	57
23	909.759	0.00014165	1	0	8	2	12	57
24	843.354	0.00014165	1	0	8	1	12	58
25	954.664	0.00014165	1	0	8	2	12	57
26	592.917	0.00014165	1	0	8	2	12	57
27	493.215	0.0001648	1	2	7	1	13	58
28	609.78	0.0001648	1	1	7	1	13	58
29	578.252	0.0001648	1	1	7	1	13	58
30	609.012	0.0001648	1	0	7	1	13	58
31	860.781	0.0001648	1	0	7	1	13	58
32	395.027	0.0001648	1	0	7	1	13	58
33	638.022	0.00020162	1	0	11	7	9	52
34	446.743	0.00028093	1	2	9	4	11	55
35	471.706	0.00028093	1	1	9	4	11	55
36	391.655	0.00028093	1	0	9	4	11	55
37	406.246	0.00041626	1	1	11	8	9	51
38	555.274	0.00044086	1	0	8	2	12	57
39	535.247	0.00044086	1	0	8	3	12	56
40	726.273	0.00044086	1	0	8	3	12	56
41	1247.101	0.00044086	1	0	8	3	12	56
42	1111.3781	0.00044086	1	0	8	3	12	56
43	1047.707	0.00044086	1	0	8	3	12	56
44	988.661	0.00044086	1	0	8	3	12	56
45	412.891	0.00044086	1	0	8	3	12	56
46	607.288	0.000625	1	2	7	2	13	57
47	909.76	0.000625	1	1	7	1	13	58
48	504.238	0.000625	1	1	7	2	13	57
49	1153.5341	0.000625	1	0	7	2	13	57
50	827.956	0.000625	1	0	7	2	13	57
51	854.378	0.000625	1	0	7	2	13	57
52	1471.514	0.000625	1	0	7	2	13	57
53	507.228	0.000625	1	0	7	1	13	58
54	838.257	0.000625	1	0	7	2	13	57
55	1234.38	0.000625	1	0	7	2	13	57
56	422.725	0.00064492	1	2	5	0	15	59
57	684.814	0.00064492	1	1	5	0	15	59
58	483.245	0.00064492	1	4	5	0	15	59
59	905.393	0.00064492	1	1	5	0	15	59
60	863.342	0.00064492	1	2	5	0	15	59
61	451.219	0.00064492	1	3	5	0	15	59
62	483.288	0.00064492	1	2	5	0	15	59
63	497.73	0.00064492	1	1	5	0	15	59
64	455.246	0.00064492	1	1	5	0	15	59
65	498.795	0.00064492	1	1	5	0	15	59
66	1222.775	0.00064492	1	1	5	0	15	59
67	425.223	0.00064492	1	2	5	0	15	59
68	456.709	0.00064492	1	1	5	0	15	59
69	668.291	0.00064492	1	1	5	0	15	59
70	557.565	0.00064492	1	0	5	0	15	59
71	1183.4871	0.00064492	1	0	5	0	15	59
72	760.738	0.00064492	1	0	5	0	15	59
73	629.717	0.00064492	1	0	5	0	15	59
74	952.725	0.00064492	1	0	5	0	15	59
75	523.807	0.00064492	1	0	5	0	15	59

	mz	p	q	Cocci Lysate	True pos	False Pos	False Neg	True Neg
	NA	NA	NA					
76	1125.407	0.00064492	1	0	5	0	15	59
77	873.895	0.00064492	1	0	5	0	15	59
78	1064.946	0.00064492	1	0	5	0	15	59
79	1064.95	0.00064492	1	0	5	0	15	59
80	500.736	0.00064492	1	0	5	0	15	59
81	816.709	0.00064492	1	0	5	0	15	59
82	497.511	0.00064492	1	0	5	0	15	59
83	492.663	0.00064492	1	0	5	0	15	59
84	628.243	0.00064492	1	0	5	0	15	59
85	819.707	0.00064492	1	0	5	0	15	59
86	529.757	0.00064492	1	0	5	0	15	59
87	756.421	0.00064492	1	0	5	0	15	59
88	602.337	0.00064492	1	0	5	0	15	59
89	474.13	0.00064492	1	0	5	0	15	59
90	565.197	0.00067194	1	0	9	5	11	54
91	568.298	0.00075648	1	4	6	1	14	58
92	746.665	0.00075648	1	1	6	1	14	58
93	713.998	0.00075648	1	2	6	1	14	58
94	431.187	0.00075648	1	1	6	1	14	58
95	806.452	0.00075648	1	1	6	1	14	58
96	456.244	0.00075648	1	2	6	0	14	59
97	1286.092	0.00075648	1	1	6	1	14	58
98	681.28	0.00075648	1	1	6	1	14	58
99	550.922	0.00075648	1	0	6	1	14	58
100	677.758	0.00075648	1	0	6	1	14	58
101	677.398	0.00075648	1	0	6	1	14	58
102	921.322	0.00075648	1	0	6	1	14	58
103	959.421	0.00075648	1	0	6	1	14	58
104	1109.0291	0.00075648	1	0	6	1	14	58
105	926.411	0.00075648	1	0	6	1	14	58
106	537.971	0.00075648	1	0	6	1	14	58
107	803.271	0.00075648	1	0	6	1	14	58
108	1216.088	0.00075648	1	0	6	1	14	58
109	697.219	0.00075648	1	0	6	1	14	58
110	991.304	0.00075648	1	0	6	1	14	58
111	504.25	0.00075648	1	0	6	1	14	58
112	1343.8	0.00075648	1	0	6	1	14	58
113	449.72	0.00075648	1	0	6	0	14	59
114	675.771	0.00075648	1	0	6	0	14	59
115	1349.804	0.00079308	1	0	10	7	10	52
116	513.726	0.0011206	1	2	8	3	12	56
117	605.764	0.0011206	1	1	8	3	12	56
118	778.331	0.0011206	1	0	8	4	12	55
119	755.929	0.0011206	1	0	8	3	12	56
120	457.262	0.00143224	1	1	9	6	11	53
121	430.886	0.00143224	1	0	9	5	11	54
122	558.249	0.00153242	1	1	10	7	10	52
123	439.19	0.00153242	1	0	10	7	10	52
124	554.262	0.0017528	1	4	7	2	13	57
125	575.287	0.0017528	1	1	7	2	13	57
126	482.225	0.0017528	1	1	7	2	13	57
127	648.294	0.0017528	1	2	7	2	13	57
128	704.835	0.0017528	1	1	7	3	13	56
129	568.745	0.0017528	1	1	7	3	13	56
130	507.264	0.0017528	1	1	7	3	13	56
131	835.845	0.0017528	1	0	7	3	13	56
132	590.004	0.0017528	1	0	7	3	13	56
133	956.348	0.0017528	1	0	7	3	13	56
134	419.314	0.0017528	1	0	7	2	13	57
135	928.408	0.0017528	1	0	7	3	13	56
136	546.581	0.0017528	1	0	7	3	13	56
137	709.854	0.0017528	1	0	7	2	13	57
138	656.29	0.0017528	1	0	7	3	13	56
139	449.19	0.0017528	1	0	7	3	13	56
140	427.707	0.0017528	1	0	7	3	13	56
141	460.915	0.0017528	1	0	7	3	13	56
142	450.757	0.0017528	1	0	7	3	13	56
143	474.694	0.0017528	1	0	7	3	13	56
144	671.308	0.00246407	1	1	8	4	12	55
145	571.754	0.00246407	1	1	8	5	12	54
146	460.212	0.00246407	1	1	8	5	12	54
147	661.799	0.00246407	1	2	8	4	12	55
148	712.259	0.00246407	1	0	8	5	12	54
149	434.88	0.00246407	1	0	8	4	12	55
150	538.324	0.00246407	1	0	8	4	12	55

APPENDIX F

TOP 150 DOG GLYCAN M/Z

	mz	p	q	TRUE pos	FLASE Neg	True Neg	FALSE Pos
	NA	NA	NA	TRUE pos	FLASE Neg	True Neg	FALSE Pos
1	468.731	2.33E-08	0.01697056	10	1	49	3
2	618.785	6.18E-07	0.44939447	10	1	46	6
3	542.249	6.18E-07	0.44939447	10	1	46	6
4	733.998	3.25E-06	1	10	1	44	8
5	580.263	1.32E-05	1	10	1	42	10
6	609.779	4.42E-05	1	10	1	40	12
7	601.276	0.00345719	1	10	1	30	22
8	626.284	0.00902491	1	10	1	27	25
9	425.863	2.32E-09	0.00168903	9	2	52	0
10	580.264	4.61E-07	0.33506478	9	2	49	3
11	609.78	2.07E-05	1	9	2	45	7
12	546.576	2.07E-05	1	9	2	45	7
13	948.539	2.07E-05	1	9	2	45	7
14	601.277	4.23E-05	1	9	2	44	8
15	726.278	4.23E-05	1	9	2	43	9
16	626.285	8.14E-05	1	9	2	43	9
17	704.264	8.14E-05	1	9	2	43	9
18	726.834	0.00111162	1	9	2	38	14
19	609.274	0.00111162	1	9	2	38	14
20	1112.4	0.00170163	1	9	2	36	16
21	618.784	0.00254126	1	9	2	36	16
22	638.289	0.00254126	1	9	2	35	17
23	755.345	0.00254126	1	9	2	36	16
24	739.328	0.00254126	1	9	2	36	16
25	632.694	0.00530935	1	9	2	34	18
26	733.997	0.01396626	1	9	2	31	21
27	933.428	0.04122187	1	9	2	26	26
28	661.803	6.04E-06	1	8	3	49	3
29	607.606	6.04E-06	1	8	3	49	3
30	819.374	1.72E-05	1	8	3	48	4
31	520.244	1.72E-05	1	8	3	48	4
32	905.92	4.24E-05	1	8	3	47	5
33	819.893	4.24E-05	1	8	3	47	5
34	1001.487	4.24E-05	1	8	3	47	5
35	1111.502	4.24E-05	1	8	3	47	5
36	726.279	9.38E-05	1	8	3	45	7
37	1132.017	9.38E-05	1	8	3	46	6
38	704.265	0.00019015	1	8	3	45	7
39	723.994	0.00019015	1	8	3	45	7
40	1100.4919	0.00035959	1	8	3	44	8
41	632.695	0.00035959	1	8	3	44	8
42	949.039	0.0006416	1	8	3	43	9
43	425.862	0.00108976	1	8	3	42	10
44	512.99	0.00108976	1	8	3	42	10
45	1108.49	0.0017742	1	8	3	42	10
46	607.605	0.0017742	1	8	3	41	11
47	620.289	0.00278403	1	8	3	41	11
48	661.802	0.00422921	1	8	3	39	13
49	1001.821	0.00422921	1	8	3	39	13
50	876.42	0.00624214	1	8	3	38	14
51	755.346	0.00624214	1	8	3	38	14
52	724.328	0.00624214	1	8	3	37	15
53	526.244	0.01261682	1	8	3	36	16
54	1020.452	0.01735826	1	8	3	35	17
55	546.575	0.0234237	1	8	3	34	18
56	739.327	0.0234237	1	8	3	34	18
57	554.246	5.96E-07	0.43351642	7	4	52	0
58	618.786	5.96E-07	0.43351642	7	4	52	0
59	546.577	5.96E-07	0.43351642	7	4	52	0
60	821.332	4.47E-06	1	7	4	51	1
61	622.601	4.47E-06	1	7	4	51	1
62	525.69	4.47E-06	1	7	4	51	1
63	1020.203	4.47E-06	1	7	4	51	1
64	733.999	4.47E-06	1	7	4	51	1
65	949.042	4.47E-06	1	7	4	51	1
66	955.529	4.47E-06	1	7	4	51	1
67	825.296	4.47E-06	1	7	4	51	1
68	413.152	4.47E-06	1	7	4	51	1
69	942.943	4.47E-06	1	7	4	51	1
70	1110.541	4.47E-06	1	7	4	51	1
71	1052.994	1.89E-05	1	7	4	50	2
72	536.26	1.89E-05	1	7	4	50	2
73	726.28	1.89E-05	1	7	4	49	3
74	819.894	1.89E-05	1	7	4	50	2
75	791.459	1.89E-05	1	7	4	50	2

	mz	p	q	TRUE pos	FLASE Neg	True Neg	FALSE Pos	In lysate?
	NA	NA	NA					
76	826.647	1.89E-05	1	7	4	50	2	0
77	847.889	1.89E-05	1	7	4	50	2	0
78	584.6	5.88E-05	1	7	4	49	3	0
79	704.276	5.88E-05	1	7	4	49	3	0
80	1080.516	5.88E-05	1	7	4	49	3	0
81	1346.105	5.88E-05	1	7	4	49	3	0
82	900.385	5.88E-05	1	7	4	49	3	0
83	819.36	0.00015112	1	7	4	48	4	2
84	690.314	0.00015112	1	7	4	47	5	2
85	876.511	0.00015112	1	7	4	48	4	1
86	1040.975	0.00015112	1	7	4	48	4	0
87	1021.459	0.00015112	1	7	4	48	4	0
88	792.854	0.00015112	1	7	4	48	4	0
89	609.275	0.00015112	1	7	4	48	4	0
90	1060.834	0.00015112	1	7	4	48	4	0
91	949.04	0.00015112	1	7	4	48	4	0
92	539.002	0.00015112	1	7	4	48	4	0
93	734.332	0.00033861	1	7	4	47	5	2
94	617.279	0.00033861	1	7	4	47	5	2
95	632.696	0.00033861	1	7	4	47	5	0
96	812.325	0.00033861	1	7	4	47	5	0
97	545.014	0.00033861	1	7	4	47	5	0
98	1075.988	0.00033861	1	7	4	47	5	0
99	650.758	0.00033861	1	7	4	47	5	0
100	876.421	0.00033861	1	7	4	47	5	0
101	1096.509	0.00033861	1	7	4	47	5	0
102	718.281	0.00068418	1	7	4	45	7	2
103	1114.509	0.00068418	1	7	4	46	6	0
104	1024.974	0.00068418	1	7	4	46	6	0
105	932.941	0.00068418	1	7	4	46	6	0
106	1041.481	0.00068418	1	7	4	46	6	0
107	833.386	0.00127461	1	7	4	45	7	1
108	626.286	0.00127461	1	7	4	45	7	1
109	745.357	0.00127461	1	7	4	45	7	0
110	904.428	0.00127461	1	7	4	45	7	0
111	997.459	0.00127461	1	7	4	45	7	0
112	1010.86	0.00127461	1	7	4	45	7	0
113	1119.538	0.00127461	1	7	4	45	7	0
114	933.431	0.00222351	1	7	4	44	8	1
115	505.731	0.00222351	1	7	4	44	8	2
116	687.32	0.00222351	1	7	4	44	8	2
117	1112.401	0.00367322	1	7	4	42	10	3
118	762.871	0.00367322	1	7	4	43	9	1
119	1183.557	0.00367322	1	7	4	44	8	0
120	833.385	0.00367322	1	7	4	43	9	0
121	618.286	0.00367322	1	7	4	43	9	0
122	1024.474	0.00367322	1	7	4	43	9	0
123	734.915	0.00367322	1	7	4	43	9	0
124	690.313	0.00579534	1	7	4	42	10	2
125	1103.504	0.00579534	1	7	4	42	10	1
126	736.005	0.00579534	1	7	4	42	10	1
127	1118.538	0.00579534	1	7	4	42	10	0
128	833.384	0.00579534	1	7	4	42	10	0
129	884.416	0.00878981	1	7	4	41	11	1
130	790.863	0.00878981	1	7	4	41	11	0
131	904.455	0.00878981	1	7	4	41	11	0
132	1020.702	0.00878981	1	7	4	41	11	0
133	542.248	0.01288224	1	7	4	40	12	1
134	717.271	0.01288224	1	7	4	40	12	0
135	1001.486	0.01288224	1	7	4	40	12	0
136	830.393	0.01831974	1	7	4	39	13	1
137	908.448	0.01831974	1	7	4	40	12	0
138	1075.984	0.01831974	1	7	4	40	12	0
139	948.538	0.01831974	1	7	4	39	13	0
140	718.662	0.02536501	1	7	4	38	14	1
141	1067.988	0.03428902	1	7	4	37	15	1
142	712.326	0.03428902	1	7	4	37	15	0
143	1112.399	0.0453624	1	7	4	36	16	2
144	723.993	0.0453624	1	7	4	36	16	0
145	734.917	6.80E-06	1	6	5	52	0	1
146	718.335	6.80E-06	1	6	5	52	0	1
147	501.25	6.80E-06	1	6	5	52	0	1
148	911.437	6.80E-06	1	6	5	52	0	1
149	607.607	6.80E-06	1	6	5	52	0	0
150	492.559	6.80E-06	1	6	5	52	0	0

APPENDIX G

MASS SPECTROMETRY METHOD PARAMETERS

	Retention [min]	Flow [µl/min]	%B	Curve	
1	-1.000	0.300	0.0		
2	0.000	0.300	0.0	5	
3	0.000	0.300	0.0	5	
4	0.100	0.300	1.0	5	
5	13.000	0.300	1.0	5	
6	85.000	0.300	37.0	5	
7	90.000	0.300	95.0	5	
8	100.000	0.300	95.0	5	
9	101.000	0.300	1.0	5	
10	120.000	0.300	1.0	5	
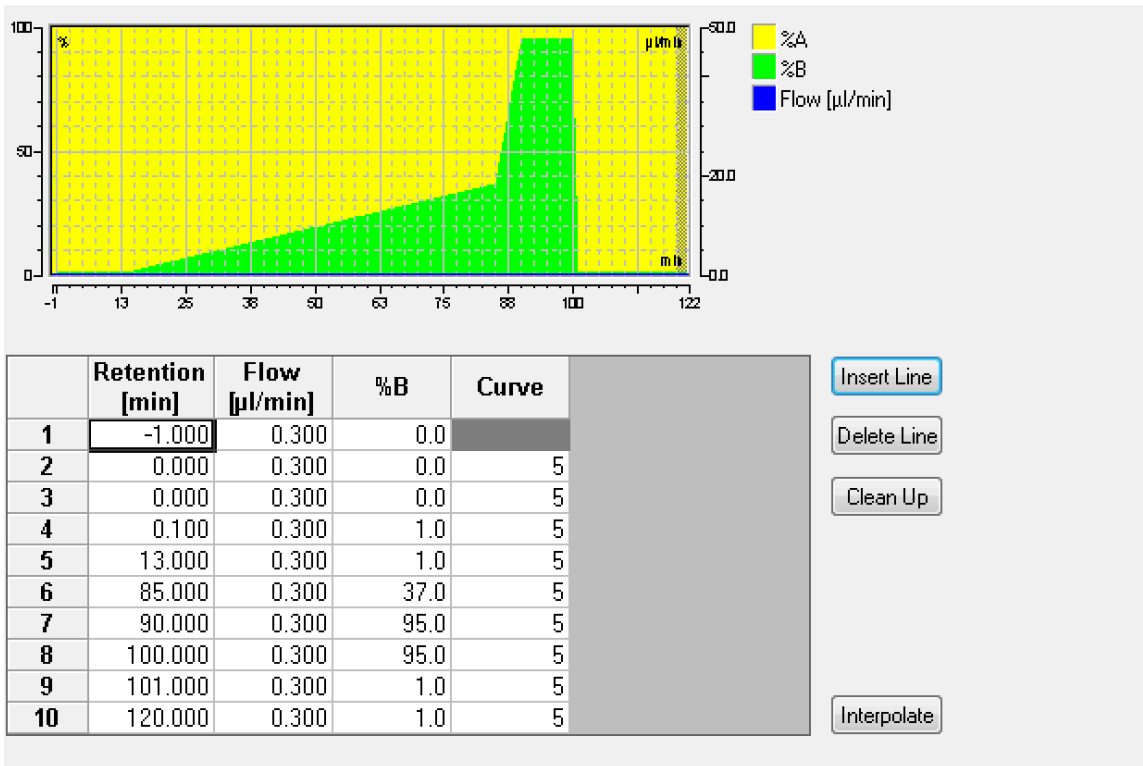
HPLC gradient curve used in this study.

MS1 scan parameters used in this study.

MS2 scan parameters used in this study.

Scan summary parameters used in this study.