1-1-2013

Development of Proteomic Characterization and Speciation Techniques Utilizing Tryptic Peptides with MALDI-TOF MS and LC-ESI MS-MS

Jennifer Marie Kooken
University of South Carolina

DEVELOPMENT OF PROTEOMIC CHARACTERIZATION AND SPECIATION TECHNIQUES UTILIZING TRYPTIC PEPTIDES WITH MALDI-TOF MS AND LC-ESI MS-MS

by

Jennifer Marie Kooken

Bachelors of Science
Syracuse University, 2006

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Biomedical Sciences

School of Medicine

University of South Carolina

2013

Accepted by:

Alvin Fox, Major Professor
Chairman, Examining Committee

Karen Fox, Committee Member

Kim Creek, Committee Member

Lucia Creek, Committee Member

Cory Robinson, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

DEDICATION

This dissertation is dedicated with all my love to

Ariel Sally

It is hard to find someone who loves, supports, trusts, and cares for you.

It is even more difficult to realize that you feel the same way towards them.

Then add in a PhD, distance, and war to make life seem impossible.

You have made the seemingly impossible possible

Thank you

ABSTRACT

The characterization of microbes which can be opportunists and pathogens (e.g., methicillin resistant *Staphylococcus aureus* (MRSA)) is important in understanding and potentially treating diseases caused by various bacterial species. Common genera found in the human skin micro-biome include *Micrococcus* and *Staphylococcus*, but there only a limited number of tests to differentiate these genera and/or species. My research reflects methods development from distinguishing one closely related genera from another and then expanded to species identification. Tryptic peptides were analyzed by MALDI TOF MS and the mass profiles compared with those of a reference strain in both genus and species identification. Aconitate hydratase and oxoglutarate dehydrogenase served as marker proteins on focused analysis after gel separation. Alternatively on full proteomics analysis elongation factor tu provided the highest confidence in staphylococcal speciation. Ultimately, refinement in speed and accuracy of analysis was accomplished with LC ESI MS-MS for peptide analysis on whole proteomes and data analysis also employing a customized X!tandem database allowing for successful identification and differentiation of the 11 Staphylococcus species commonly found on humans (genomes gathered from UNIPROT). The methodological approach described can be utilized for bacterial identification across multiple species,

creating more accurate methods for identification of species that may be pathogenic to humans.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.0 ***Development of Dissertation***

In the field of bacteriology, proteomics is largely applied to cataloging protein expression, with little emphasis on taxonomic identification. In Proteomics the main concern of analyzing proteins in a given cell or organism (referred to as its proteome) is quantifying the abundance and changes that occur over time with these proteins (Phillips and Bogyo, 2005). This work focuses on identifying coagulase negative Staphylococcus (CoNS) species and to introduce methods in which mass spectrometry (MS) may be used for bacterial species identification. The need for improved taxonomic identification stems from current methodology failing in proper speciation of strains. Most CoNS species were believed to hold no clinical significance, but it is now being found CoNS can be found in clinical settings. Furthermore, even clinically significant Staphylococcus species are often misidentified creating a need for consistently accurate taxonomic classification method. Current physiological and molecular tests often fail in proper identification of CoNS. Better, although not complete accuracy in species identification has been found with utilization of mass spectrometry, such as matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) (Fox et al., 2011).

What will be addressed herein are the current problems with taxonomic identification of CoNS, and the need to establish a methodology for accurate identification of both genus and species from samples across multiple origins. Along with using *Staphylococcus* as a model system for bacterial speciation, microbial characterization is important in understanding transmission (Fox et al., 2008). Bacterial transmission may occur person-to-person, or there is potential for its presence from environmental sources (Fox et al., 2008). Methodology for species identification of common organisms causing human disease is widely available in clinical microbiology laboratories (Fox et al., 2008). Even with this ability new and closely related disease causing species often fail to be characterized, or are miscatagorized, in current clinical settings (Fox et al., 2011). This research was performed focusing on 11 species of *Staphylococcus* commonly identified in clinical settings, with many of these species often misidentified (Goyal R, 2006; Surekha.Y.Asangi 2011). The progression of this dissertation establishes a methodology to identify the bacterial Genus of *Staphylococcus*, and further explore accurate identification utilizing bacterial proteins for species identification.

1.1 **General approaches to taxonomy**

Microbiological testing is often used in clinical settings for bacterial species identification. The ability to culture and stain bacterial samples allows basic information to be attained. Gram staining is a technique performed to ascertain if the sample contains a thick layer of peptidoglycan. Peptidoglycan will stain purple if it is gram

positive. Conversely, if this layer is lacking and the sample instead stains red (gram negative). Physiological tests such as evaluation of hemolytic activity, ability to ferment various sugars, along with antibiotic resistance can be utilized to eliminate possibilities based upon known properties of bacteria. Many of these tests have assisted in species identification. Although often the results may be inconclusive when problems such as partial gram staining, or incomplete fermentation of sugars lead to unreliable tests. This creates difficulties that are both time consuming and lead to the inaccuracies in species identification.

With physiological tests leaving many inconclusive results microbiological methods has become the standard for species identification. Tests such as PCR and 16s rRNA allow for a more rapid identification then previous physiological methods, but still have many drawbacks. With Polymerase Chain Reactions (PCR) a specific region of DNA is amplified. This region should be specific to the bacteria targeted and when performed identification is quite often successful especially when there is greater divergence among species in the variable regions of these genes; the problem that often arises with using this method is species identification (Goebel and Stackebrandt, 1994; Goebel, 1994). The genetic region selected for PCR analysis must contain two conserved regions for primers to recognize and allow amplification to begin. It is believed that the variable sequence between these conserved regions allow for taxonomic discrimination, but this also can become the problem with accurate identification. Similar to most PCR based techniques the requirement for two conserved regions limits the number of genetic regions available to study. Sequence variation of genes, such as sodA may be so subtle that differentiation of closely related species

cannot be obtained. Also, in most cases this would require some knowledge of the sample in order to create the appropriate primer for the PCR process to occur. The dependence on this prior information to properly create primers that identify species this may not be a suitable technique in identification of unknown bacterial isolates. Accurate results of species identification does not occur when species are closely related but unsequenced. Microarrays also utilize genetic information similar to PCR, with the same drawbacks from closely related species ability to cross hybridize on arrays giving inaccurate results (Jukes et al., 2010).

Another popular molecular technique is 16S ribosomal DNA (rDNA) sequencing, which is also used as the gold standard in microbial ecology and serves as a cataloging technique (Fox et al., 2011). Overall it was discovered that 16S rDNA sequencing discriminates well at the genus level, with most clinically significant species cataloged, but species-level identification capabilities often fall short (Morgan et al., 2009). Problems arise due to 16S ribosomal DNA sequences having closely related species that contain identical 16S rDNA sequences or, alternatively, that divergent 16S rDNA sequences may exist within a single organism (Goebel, 1994). This situation was observed in a study of airborne urban dust where 16S rRNA arrays indicated the presence of over 8,000 taxa, and none were identifiable to the species level (Brodie et al., 2007). With physiological and molecular techniques able to narrow down unknown samples to their respective genus, there is still a failure in obtaining consistently correct and accurate species level identification. This problem, being well recognized the field of proteomics, is being explored to fill in the gaps from genomic techniques.

1.2 **Approaches to *Staphylococcal* Taxonomy**

Taxonomically, the genus *Staphylococcus* is in the bacterial family Staphylococcaceae, which is also comprised of Gamella, Macrococcus and Salinicoccus(Baird-Parker, 1965). The genus *Bacillus* in the family Bacillaceae is the most well-known phylogenetic relative (Baird-Parker, 1965). *Staphylococci* are gram-positive bacteria, spherical in shape and under a microscope closely resemble grape clusters (Baird-Parker, 1965).

Identification of *Staphylococcus* species have several complications, along with ambiguous gram staining results, glucose and glycerol tests are often inconclusive leaving a potentially pathogenic *Staphylococcus* sample to be misidentified for the morphologically similar, yet harmless species of *Micrococcus* (Baker, 1984). It has been found that in human cultures of the nose and skin results will often produce staphylococci (Roth and James, 1988). From these cultures clinical identification systems such as the STAPH-IDENT system utilize multiple tests to identify *Staphylococcus* samples and further specieate these samples. In a four year study it was found these test strips did not perform adequately when dealing with commonly encountered members of the family Micrococcaceae; and was unsuitable for the identification of uncommon *Staphylococcus* isolates (Rhoden and Miller, 1995). Currently there are more than 20 species of *Staphylococcus* described in Bergey's Manual (2001), but only *Staphylococcus aureus* and *Staphylococcus epidermidis* are considered significant in their interactions with humans. This outlook is slowly changing as many other species of *Staphylococcus* are now appearing in human infections

(Dubois et al., 2010). The frequency in which other CoNS species associated with human disease is reported varies from study to study. With the vast range of reporting it is expected that clinical labs are often misidentifying species creating the problem of unknown prevalence in CoNS infections. It is reported that *S. hominis, S. warneri, S. capitis, S. haemolyticus and S. lugdenensis* are the next most common species in staphylococcal infections, with *S. simulans* rarely reported from clinical microbiology laboratories (Sivadon et al., 2005). In contrast, studies of prosthetic joint infections found *S. lugdenensis* and *S. simulans* (along with *S. capitis*) as being the major organisms isolated (Frank et al., 2004). The variation in species reported, and lack of a standard method in which to identify a species, creates a situation where misidentification of disease causing organisms may lead to improper treatments. Since realization that there are many species of coagulase negative *Staphylococcus* that also inhabit the human skin, and physiological tests are unable to conclusively identify species PCR was then used to find proper identification (Sivadon et al., 2005; Skow et al., 2005). The highly conserved hsp60 and tuf genes have been found to be useful for identification and taxonomic classification along with sodA genes in *Staphylococcus* species identification (Goebel, 1994).

Staphylococci remain an important component of the human flora and laboratories normally equated coagulase-positive, gram-positive cocci with *Staphylococcus aureus* and coagulase-negative cocci with *Staphylococcus epidermidis*, but there are at least 13 human strains of coagulase-negative *staphylococci* that are also pathogenic and often misidentified (Rhoden and Miller, 1995). From the 13 species commonly found on humans eleven are readily available from the ATCC and

have complete or close to complete genomic sequence annotated within the NCBI or UNIPROT system (Kooken, 2013) In development of this dissertation the ability to identify isolates from environmental, human, and veterinary origins were accessed, with the goal of proper and consistent identification of coagulase negative *staphylococci*. Tandem mass spectrometry was utilized to explore the possibility of applying current proteomic technology to protein identification for speciation of unknown *Staphylococcus* sample.

1.3 **Use of Mass Spectrometry in Proteomics**

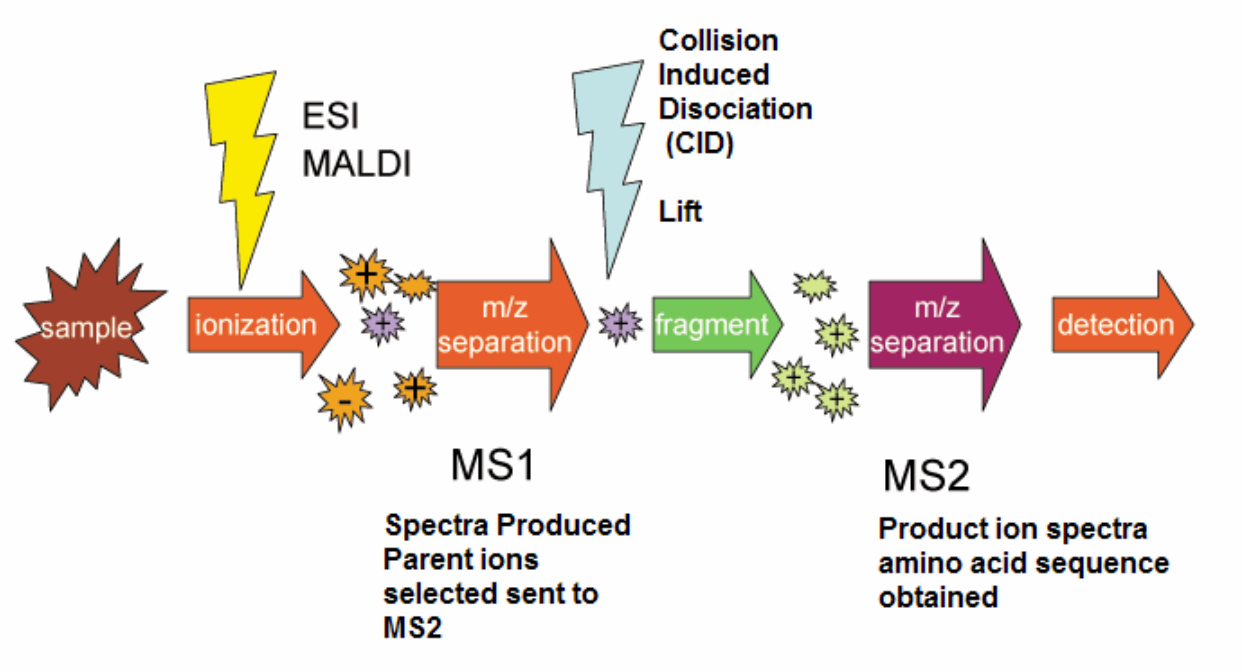
Proteomics is one of the fastest growing areas in biological research, the current focus in this field has been protein identification, with little focus on how this information may be utilized in species determination (Aebersold and Mann, 2003). The ability to discriminate bacterial isolates is important for many areas of research in medical microbiology, especially in defining bacterial taxonomy and monitoring transmission of infection (Cash, 2009). Proteins analysis allows us an understanding of the functions and regulation occurring in an organism, giving a picture of current activity occurring in the cell, and what proteins are being produced (Intelicato-Young and Fox, 2013). In the past genomic studies provided a vast amount of overall information but failed in identifying what is occurring at any given time point; which illustrates an incomplete picture of the organism (Intelicato-Young and Fox, 2013). In comparison, the proteome, which is complementary to the genome, is able to display a broader picture of biological occurrences within a cell (Intelicato-Young and Fox, 2013). The genome is a depiction

of everything that could possibly occur within an organism, but the proteome shows what is currently active within the cell (Intelicato-Young and Fox, 2013). Proteomic analysis allows for comprehensive identification of proteins, along with valuable information such as prevalence of a protein in a given sample, isoforms, and helps to identify functions of each protein (Angel et al., 2012).

One of the most valuable technological advances in proteomic study is mass spectrometric analysis. Mass spectrometry (MS) is a widely utilized analytical technique that has been refined over the years to its current state. Sensitivity and selectivity of this technique allow us to gather information on a compounds molecular weight (MW) and even structural information in a short time. The potential of these advances outside of chemistry was not fully realized until the early 1990ôs when MS became widely applied to study peptides and proteins (Canas et al., 2006). The ability to contribute to protein identification did not occur until Electrospray ionization (ESI) and Matrix-Assisted Laser Desorption/Ionization (MALDI) was developed and large quantities of peptides and proteins could be identified from small quantities of material (Fenn et al., 1989; Whitehouse et al., 1985). Karas and Hillenkamp published spectra of large proteins obtained by coupling their MALDI source to a time-of-flight mass spectrometer (MALDI-TOF MS) (Karas and Hillenkamp, 1988). This development assisted in ushering in the age of proteomics. The information gathered from mass spectrometry, coupled with computer software that compares experimental versus theoretical spectra allowed fingerprints to be established with spectral patterns. These fingerprints were observed to be unique enough to identify a protein.

Utilizing these approaches it is possible for characterization, quantitation, and even identification of unknown protein samples (Aebersold and Mann, 2003). With standard MS techniques further coupled for tandem MS the available genomic research can be utilized in identifying unique peptides within the protein for species identification. Tandem mass spectrometry allows direct protein analysis providing accurate amino acid sequence without the need for any prior information.

1.4.0 **Mass Spectrometer overview**



Slide courtesy of K. Murray, altered for dissertation purposes

Figure 1.1 General overview of MS and MS-MS separation

Mass spectrometry is a tool that accurately measures the mass of different molecules within a sample. With MS it is possible to identify molecules in a mixture, detect impurities in a sample, analyze a purified protein, and as with this research to study the protein content in cells (Baker, 2010). There are 3 main aspects to a mass spectrometer; ionization, sorting, and detection. Ionization is where the molecules in a sample are converted to the gas phase. This takes place by heating allowing the subsequent vapors to be ionized by bombardment with an electron rich beam. Which in turn creates ions of both positive and negative charges. Next these ions are sorted

according to mass, this sorting may occur through acceleration, deflection, or a combination of the two dependent upon the MS system used. The last part of MS would be detection; upon analysis of the ions that reach the detector the computer will generate a spectrum representative of the ions found. This spectrum can serve as the fingerprint for the sample and applied towards identification. The electrical or magnetic fields that are generated from the mass spectrometer is inversely proportional to the mass, and directly proportional to its electrical charge; so it is indicated as a mass to charge ratio (m/z) (Hillenkamp et al., 1991; Karas and Hillenkamp, 1988). The spectrum readout is displayed in this fashion so dependent upon molecule charge it is possible to ascertain the exact mass. Tandem MS adds a secondary layer of information. Parent ions from the MS spectra may be selected and subjected to a second fragmentation which breaks the peptide into its amino acid components.

1.4.1 **Mass Spectrometer Ion sources**

Conditional for the type of sample analyzed there are various ionization techniques utilized in mass spec. With proteomics these biological samples are normally looked at with soft ionization. Opposed to hard ionization where chemical bonds are broken, soft ionization results in the formation of ions without breaking any chemical bonds. Soft ionization techniques allow for all covalent bonds to be kept intact. This method of ionization is also preferred with large biomolecules. In comparison hard techniques require the analyte to be vaporized first then generate ions from neutral molecules that are now in the gaseous phase, causing breaking of the

molecule (Tanaka et al., 1988). Hard ionization is not utilized because proteomic samples normally are polar and non-polar and are more readily dissolved into an aqueous solution. Heating theses samples to place them in a gaseous phase would only succeed to cook the sample instead of vaporizing. Soft ionization allows these fragile molecules to enter the gas phase without using so much energy that they fall apart, allowing the ability to measure mass.

The most common soft ionization sources, in proteomic work are electrospray (ESI) and MALDI (Angel et al., 2012). Mass analyzers utilized in proteomics have a larger variety; ion traps, triple quadropole; Fourier transform cyclotrons; and time-of-flight (TOF) are ones frequently used (Angel et al., 2012). All of these analysis methods are based upon achieving the mass to charge separation in an electromagnetic field. If a molecule is neutral it needs to first be converted into a charged ion. In proteomics the size of proteins and protein peptides have to be taken into consideration when deciding upon ion source, ESI and MALDI have both proven to generate peptide and protein ions efficiently even for molecules over 10,000 Da (Van Riper et al., 2013).

Even with success in proteomic analysis both ion sources have benefits and drawbacks. With ESI, ion formation occurs at atmospheric pressure (Laiko et al., 2000a; Laiko et al., 2000b). A high voltage is applied as the liquid is passed through a heated capillary or nitrogen depending on the type of ESI source (Whitehouse et al., 1985). These droplets are so small that they can evaporate rapidly concentrating the charged molecules and acquire multiple charges in this process. This allows the peptides or proteins to be multiply-protonated, causing various mass to charge ratios that can enhance ion transmission in the subsequent analyzer. ESI sources can be

coupled with an HPLC to reduce volumes used even further (up to nanoliter volumes) and HPLC also provides a further means of separation to achieve greater sensitivity during analysis (Davis et al., 1995). Drawbacks to ESI are primarily due to it being a 'soft ionization' technique, there is little fragmentation, which is beneficial because the molecular ions are often repeatedly observed across multiple sample runs, but it gives very little structural information unless further tandem MS is performed.

An often used alternative to ESI ion generation is MALDI. In MALDI ions may be generated either at atmospheric pressure, but more commonly under vacuum conditions (Laiko et al., 2000b). Unlike ESI, MALDI is based on a matrix, frequently Ŭ-cyano-4-hydroxycinnamic acids are used with protein and peptide analysis, due to their ability to absorb UV light (Billeci and Stults, 1993; Henzel et al., 1993). The peptide or proteins are pipetted onto the MALDI plate allowing the sample to evaporate and co-crystallize with the matrix. The plateŝs samples are then placed into a vacuum chamber and hit by a pulsed UV laser beam (Knochenmuss, 2006). The lasers energy is absorbed by the matrix and vaporizes small charged molecules allowing them to enter gaseous phase that can later be detected by the analyzer (Knochenmuss, 2006). MALDI is frequently paired with a TOF analyzer, but other methods have been used to analyze ions produced from MALDI sources. This is another soft ionization method that is successful in fragmenting large and small biomolecules such as peptides and proteins, but sample preparation is important for sensitivity, reproducibility, and quantification of mass analysis (Carbonnelle et al., 2012; Carbonnelle et al., 2011). MALDI was the first MS technique that had successful protein and polymer analysis up to m/z 100,000 (Tanaka et al., 1988). But due to the sample preparation methods the

protein/peptide-matrix mixture is not homogeneous. It is necessary to take multiple laser shots at different places on the sample and results shown are the statistical average of the substance concentration in that MALDI sample and can subsequently loose information on smaller peptides (Angel et al., 2012). After either ESI or MALDI generate ions an analyzer and detector are needed to separate the ions and interpret the data into a mass spectrum.

1.4.2 **Mass Spectrometry Analyzers/Detectors**

The analyzer separates ions obtained at the ion source and applies either an electric or magnetic field to induce separation. Similar to ion sources there are many analyzers and detectors that may be used in proteomic work, but due to low price and ease of use quadrupoles, TOF, and ion traps, are the most frequently utilized (Yates, 1998). Within my research the detector aspect although important for mass spectrometry is not focused on, this is because all modifications for MS and MS-MS output occur within the source and analyzer. This allows the detector to interpret the signals received and displays these signals as a mass spectrum. Selection of an analyzer should depend upon the resolution needed and its capability to differentiate two close signals from each other; the desired standard for proteomic work is an analyzer capable of isotopic peak differentiation. The ability of a mass spectrometer to distinguish different isotopes is one of the reasons why mass spectrometry is such a powerful technique; isotopes gives each fragment a characteristic series of peaks with different intensities (Skoog, 2007). Mass spectrometry distinguishes molecules based

on their mass to charge ratio, because of this isotopes play an important role in mass spectra and each isotope will show up as a separate line in a mass spectrum with high enough resolution. If resolution is not sensitive enough then peptides with similar m/z ratios may not be distinguished and improper identification may occur. The Quadrupoles is an analyzer often used in proteomics that consist of four cylindrical rods that lay parallel to each other and exert alternating current (Edmond deHoffmann, 2003). These electrical fields oscillate and only ions with a stable trajectory will travel along the quadrupoles to reach the detector (Edmond deHoffmann, 2003). Dependent on the setting of these rods it is possible to select for ions with higher m/z ratio, all ions, or for it to allow ions with different m/z ratio to pass sequentially (Leary and Schmidt, 1996). It is also able to set fixed voltages to permit stable trajectories for specifically selected ions, this works best when the user knows the m/z ratio desired and corresponding voltages (Leary and Schmidt, 1996). The main drawback for quadropole analyzers are the mass range they are effective for, usually restricted to a limit of m/z 4000 (Chowdhury et al., 1990b). Since quadrupoles are frequently coupled with an ESI source for proteomics they are still effective in determining the molecular weight (MW) of lager proteins due to ESI ability for multiply charged ions (Chowdhury et al., 1990a).

ESI is most often paired with a quadropole mass analyzer while proteomic runs with MALDI ion sources tend to use the Time of Flight mass analyzer (TOF or TOF MS). TOF is considered one of the simplest methods for ion analysis; it consists of a flight tube in high vacuum atmosphere where ions are accelerated with equal an equal amount of energy then in a manner inversely proportional to their mass (smaller ions travel faster to the detector) they accelerate down the flight tube (Wollnik, 1993).

The spectrum produced at the end is indicative of the signal produced by ions reaching the detector at the end of the flight tube. MALDI-TOF spectrum is based on singly charged artifacts, but the peaks can be too broad to accurately identify a mass value, this happens due to the occurrence of uneven energy distributions that can degrade resolution (Jensen et al., 1996; Shevchenko et al., 1996). This problem is averted by a 100-200 nanosecond (ns) delay in applying the gradient extraction voltage, this procedure is more commonly known as delayed extraction (DE), the results may be further enhanced by utilizing an electrostatic mirror (reflectron) (Jensen et al., 1996). Often referred to as órefflectron modeô the ions pathway is reversed and travel time lengthened with this electrostatic mirror located at the end of the primary flight path (Cornish and Cotter, 1993). Due to the affordability, sensitivity, and ease of use the MALDI-TOF is a very common and highly utilized instrument in proteomics.

Thermo fisherôs Orbitrap, a replacement for the complex but high resolution Fourier transform ion cyclotron resonance (FITCR), and ion traps are the two most commonly used in proteomics. Ion traps hold ions in a fixed area allowing the analyst time to perform scans, selection of ions, fragmentation, and even product ion analysis. By retaining the ion successive analysis may be performed and repeated over the course or seconds or even hours (Blaum, 2006). There is a small space in the center of the device where ion trapping occurs, once inside the trap, increasing voltage causes ions to destabilize successively ejecting them from the trap (Stafford Jr et al., 1984). The ejected ions reach the detector in order of ejection and their m/z ratios are measured (March, 1997). Ion trap resolution is dependent upon the scanning speed, in proteomics full scans along with product scans are optimal at medium resolutions which

are needed for ion charge assignment (Hu et al., 2005). Lower scan speeds may be needed for more accurate measurements on selected smaller mass ranges, short slow scans are only compatible with high resolution instruments, such as the Orbitrap (Hu et al., 2005; Marina et al., 1999). All of these MS techniques are limited in the information they are capable of providing, but when placed in tandem (MS-MS) identification down to amino acid composition can be achieved.

1.5 **Tandem Mass Spectrometry**

Tandem MS can consist of several sequential processes; ionization of sample molecules, mass selection of parent ions, Collision induced dissociation (CID) of the parent ions with neutral gas molecules to produce daughter ions, and mass analysis and detection of daughter ions (Johnson et al., 1990). There are some alternatives to this process, such as Bruckerôs MADLI-TOF MS-MS which uses a ñLIFTò instead of CID. The LIFT portion is laser induced dissociation, which is supposed to react very similarly to CID (Suckau et al., 2003). These separation techniques usually occur in a ótandem in spaceô or ótandem in timeô configuration. With tandem in space MS-MS the separation of elements are physically distinct, where one MS apparatus, although physically connected to maintain a high vacuum atmosphere is distinct from another MS system, TOF-TOF MS-MS would be an example of separation over a time span (Johnson et al., 1990). Tandem in time MS-MS utilizes a system where the ions are trapped in one location and separation occurs in multiple steps over time, ion traps are examples of MS machines capable of performing separation over time MS (Johnson et

al., 1990). Mass spectrometry has become the foremost method for the characterization and sequencing of proteins and peptides within the field of proteomics (Shevchenko et al., 2000). Proteomic research performed for this dissertation serves as an exercise in application of modern time mass spectrometry. Utilization of MS and MS-MS coupled instruments allowed taxonomic identification down to species level across a wide range of *Staphylococcus* samples by identifying the amino acid sequences found in abundant proteins. This serves as an example why tandem mass spectrometry has become the method of choice for high resolution protein detection and identification (Cravatt et al., 2007).

1.6 General Proteomic Sample Preparations for Mass Spectrometry

The strategies used for preparation of proteins or more complex proteomic samples in MS analysis involve multiple steps. The onset of this research required bacterial culture followed by cell lysis and whole supernatant to be run on an SDS-PAGE gel. This allowed for all of the proteins to be visualized, and the 100 kD band, identified in previous research, to be excised for digestion and MALDI-TOF analysis (Fox et al., 2011). In utilizing higher resolution technology SDS-PAGE gels were eliminated and whole cell protein digests were suitable for LC-ESI-MS-MS analysis. In both circumstances the proteins are broken up into peptides, via enzymatic digestion with Trypsin. Trypsin digestion cleaves proteins at the carboxyl side of the amino acids lysine (K) and arginine (R), creating smaller, more manageable fragments for MS analysis. After Trypsin digestion particular attention is paid to the preparation of the

peptides for compatibility with the mass spectrometer. A step to clean up the sample and desalt the final peptide mixture prior to MS analysis is required, the use of spin columns, zip tips, or Solid Phase Extraction (SPE) columns all assist in this cleanup process and are based upon utilization of C18 to remove interfering salts and debris and release peptides in MS-compatible solution. These individual peptides and the information gathered from them are subsequently complied and allow analysis of the spectrum to indicate the protein identity.

1.7 **Dissertation applications**

Proteomic research performed for this dissertation served as an exercise in application of modern time mass spectrometry for identification of unknown *Staphylococcus* isolates. Utilization of MS and MS-MS coupled instruments allowed identification down to species level across a wide range of *Staphylococcal* samples. This serves as an example why MS has become the method of choice for high resolution protein detection and identification (Cravatt et al., 2007). MS and MS-MS techniques also prove to have an advantage to techniques such as cloning, PCR (polymerase chain amplification), DNA-DNA hybridization, and microarrays which are used to isolate, amplify, and identify genes in search of species identification. The following chapters will progressively establish means of differentiating *Micrococcus* from *Staphylococcus,* both of which are found on human skin and difficult to differentiate with phenotypic techniques, thus establishing genus differentiation with MALDI-TOF MS. Then expanding focus onto *Staphylococcus* speciation with specific proteins analyzed

to determine species, then overall whole cell supernatant as a faster way to establish species identity. This dissertation utilized *Staphylococcus* as the bacterial species of interest, but methods developed can be applied to a broad range of bacteria utilizing the same methods and techniques.

CHAPTER 2

Characterization of Micrococcus strains isolated from indoor air

Jennifer M. Kooken, Karen F. Fox, Alvin Fox*

Department of Pathology, Microbiology and Immunology, School of Medicine, University of South Carolina, Columbia, SC 29208, USA

* Corresponding author. Tel. 803 216 3409; fax 803 216 3428.

E-mail: alvin.fox@uscmed.sc.edu.

2.0 **Abstract**

The characterization of microbes, such as of opportunists and pathogens (e.g. methicillin resistant *Staphylococcus aureus* [MRSA]), in indoor air is important for understanding disease transmission from person-to-person. Common genera found in the human skin microbiome include *Micrococcus* and *Staphylococcus*, but there only a limited number of tests to differentiate these genera and/or species. Both genera are believed to be released into indoor air from the shedding of human skin and are morphologically difficult to distinguish. In the current work, after the extraction of proteins from *micrococci* and the separation of these proteins on one dimensional electrophoretic gels, tryptic peptides were analyzed by MALDI TOF MS and the mass profiles compared with those of a reference strain (ATCC 4698). The results confirmed that all strains were consistent in identity with *Micrococcus luteus*.

2.1 **Introduction**

In indoor air, bacteria are derived from building occupants and the environment. The characterization of airborne bacteria is important for understanding disease transmission from person-to-person. Unfortunately, building monitoring is generally limited to total colony forming units (CFUs). It is extremely time-consuming for all of the diverse colony types in indoor air to be characterized by conventional means and is rarely undertaken. Common genera found in the human skin microbiome include *Staphylococcus* and *Micrococcus* (Fox et al., 2010; Fox et al., 2011); these organisms are believed to be released into indoor air (Tang, 2003). *Staphylococcus* and *Micrococcus* are difficult to distinguish morphologically (both being observed as tetrads). Clinical samples have been reported to rarely contain micrococci (~ 4% of isolates are micrococci and 96% staphylococci) (Baker, 1984). However, two thirds of the environmental isolates in studies by us and others were *micrococci* (Satta et al., 1993; Scherer and Brown, 1974). Thus, for environmental samples there is a greater need for the characterization of *micrococci*. Of the five known *Micrococcus* species, *M. luteus* is the only one whose primary habitat is human skin (Evans and Kloos, 1972; Falk and Guering, 1983; Hebert et al., 1988); *M. lylae* is only occasionally isolated from this organ (Hebert et al., 1988).

Previously, we developed an approach for discriminating *Micrococcus* from *Staphylococcus* using MALDI TOF MS (matrix assisted laser desorption time-of-flight mass spectrometry), applying the Bruker Biotyper software and reference library (Bruker Daltonics, Bremen, Germany) and standard physiological tests. MALDI TOF MS for the

analysis of bacterial proteins from bacteria isolated from clinical samples has shown 65.7% to 98.8% accuracy for the identification to the genus level, and 31.8% to 94.2% to species (Baldellon and Megraud, 1985). In MALDI TOF MS, mass profiles are produced from isolated bacterial colonies. Colonies are sampled from bacterial culture plates, dried directly on a MALDI plate with the ionization matrix and then subjected to MALDI TOF MS analysis. The resultant protein profiles are not attributed to known proteins which complicate extrapolation of information from one laboratory to another.

Users alerted the ATCC that ATCC 9341 displayed characteristics that were quite distinct from another *M. luteus* reference strain (ATCC 4698) (Schumann et al., 2009). This encouraged the ATCC and collaborators to perform a study to characterize these two strains involving 16S rDNA sequencing. Indeed, ATCC 4698 was shown to be *Micrococcus luteus*, whereas ATCC 9341 had been designated to be a different genus (*Kokuria)* within the *Micrococcaceae* family (Schumann et al., 2009). This encouraged us to further characterize our environmental isolates and use ATCC 4698 as the reference strain. Since members of the *Micrococcaceae* are not considered to be significant human pathogens, only a limited number of strains are found in current databases.

Isolation of proteins by gel electrophoresis allows one to focus on specific protein bands and compare the relatedness of proteins from different strains by MALDI TOF MS. Additionally, MALDI TOF-TOF MS-MS allows for peptide sequencing and identification of the protein (Kocur and Martinec, 1967; Rhoden and Miller, 1995). This

is not the case for the more widely used ñdirectòMALDI TOF MS profiling (e.g., the Bruker system). Homologous proteins in closely related bacteria can provide information on species identification (e.g., heat shock proteins (Stackebrant, 1995), ribosomal proteins (Noble, 1984), and outer membrane proteins (Wieser et al., 2002).

With the continuing expansion in genomic datasets of bacteria it has been possible to identify an increasingly larger number of microorganisms. Utilization of MALDI TOF-TOF MS-MS sequencing of ions generated from tryptic digests of proteins provides the peptide sequence for comparison. This approach does not require any previous knowledge of the bacterial species. In this paper, tryptic peptides analyzed by MALDI TOF MS profiles allowed the characterization of the species identity of environmental isolates. MALDI TOF-TOF MS-MS was used to identify proteins from these peptides. Protein identification definitively categorized *Micrococcus* from environmental isolates and also distinguished them from staphylococcal species isolated from the same environment.

2.2 **Methods**

2.2.1. ***Air sampling***

Air samples were collected in occupied and unoccupied rooms in a suburban elementary school in Columbia, SC, using an N6 Single Stage, Viable Impactor (Thermo Fisher Scientific, Inc., Waltham, MA), which has a nominal flow rate of 28.3 l/min. Sheep blood agar [SBA] plates were used with the viable impactor and

subsequently incubated at 37 ęC for 24-48 h. Individual ß-hemolytic colonies were re-streaked several times until pure cultures were obtained. Gram-positive cocci occurring in tetrads or clusters are predominantly *staphylococci* and *micrococci*, and were selected for further characterization (Tang, 2003).

2.*2.2*. ***Strains***

ASO3 C5, ASO3 C6, ASO3 C10, ASO3 C15, ASO3 C17, ASO3 C24, ASO3 C45, ASO3 C46, ASO3 C55, ASO3 C68, ASUNO5 2W, ASUNO5 3W, ASUNO15 C10, ASO15 C31, ASUNO2-15, C1White, ASUNO2-15 C3Y, ASUNO2-15 C4Y, ASUNO2-15 C5Y, ASUNO2-15 C7Y, ASUNO2-15 C9Y, ASUNO2-15 C10Y, ASUNO2-15 C12Y Reference strains: *M. luteus* ATCC 4698, *M. luteus* ATCC 49732, *S. aureus* ATCC 31240, *S. hominis* ATCC 27844. *S. warneri* (ATCC 49454).

2.2.3. ***Bacterial culturing and protein extraction***

Bacteria were grown on nutrient agar and incubated at 37 °C for 24 h. Colonies were removed from plates and placed in 2 ml sterile, screw-top microcentrifuge tubes with 1 ml of protein extraction buffer (0.1 M NaCl, 50 mM Tris HCl, 0.5 mM phenylmethylsulfonyl fluoride [PMSF]). The microcentrifuge tubes were weighed before and after bacteria were added to determine total wet weight of the bacteria processed.

Samples were placed in a FastPrep®-24 (MP Biomedicals, Solon, OH) for 6 ms X 30 sec with 5 min on ice between each cycle for a total of 6 cycles. The samples were then centrifuged at 4 °C for 1 h at 10,000 x*g*. The supernatant was removed and placed at -70 °C for two freeze-thaw cycles to eliminate DNA. The supernatant containing the total protein extract was normalized to 100 mg/ml wet weight of bacteria and stored at -70 °C until used.

2.2.4. Fermentation of glucose and glycerol

Purple Agar base (Difco Manual) was prepared and 0.5% glucose or glycerol was added. The media was autoclaved for 15 min. Glycerol media was poured into petri dishes and glucose media was placed in 5 ml tubes. For testing, bacteria were streaked from Nutrient agar after 24 h of growth on to glycerol plates and incubated at 37 °C for up to 72 h. Plates were examined every 24 h for acid production. Glucose tubes were inoculated from Nutrient agar after 24 h of growth and the media covered with sterile mineral oil. The tubes were incubated at 37 °C for up to 72 h. Tubes were examined for acid production every 24 h. Previous testing revealed three *Micrococcus* strains to have weak positive fermentation patterns (glycerol one; glucose 2 others). This finding was confirmed for two of the strains (ASO3-C10 and ASO3-C17; both were weakly positive for glucose metabolism, which is inconsistent with their being *M. luteus*.

However, this was not confirmed for the other strain (AS03-C6) and does illustrate the subjectivity of reading color reactions.

2.2.5. ***Protein separation***

Prior to electrophoresis samples were vortexed, and 20 εl of supernatant were added to 20 εl of 2x loading buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M tris-HCl), and the sample was boiled at 100 °C for 5 min and centrifuged for 5 s. Samples were subjected to electrophoresis in 5% Criterion Tris-HCl gels (Bio-Rad) until the bromophenol blue was at the very bottom of the gel. The running buffer consisted of 192 mM glycine, 25 mM Tris and 0.1% SDS. The gels were washed 3 times in double de-ionized water for 5 min each and stained with Gel-Code Blue Stain reagent (Pierce, Rockford, IL) for 1 h. Gels were stored covered in double de-ionized water at 4 °C until the bands were excised.

2.2.6. ***Peptide preparation***

Bands of interest were excised from Coomassie blue-stained gels and de-stained with 25 mM ammonium bicarbonate in 50% acetonitrile/50% water. The gel was dehydrated by addition of absolute acetonitrile (covering each gel piece) for 5 min. The acetonitrile was removed and the gel rehydrated with 100 mM ammonium bicarbonate

(10 min). Then, the gel was dehydrated with acetonitrile (5 min), the acetonitrile removed and the sample dried in a Speed-Vac for 2-3 min. The sample was rehydrated with 10 mM dithiothreitol (DTT) and incubated for 30 min at room temperature (22-24 °C), then alkylated with 50 mM iodoacetamide for 30 min at room temperature, and then washed in 100 mM ammonium bicarbonate. Once the iodoacetamide had been removed, the sample went through 3 further rounds of dehydration-rehydration (100% acetonitrile- 100 mM ammonium bicarbonate). After the final dehydration, the sample was dried in a Speed-Vac. The gel spot was rehydrated (with a 12.5 ng/ml solution of trypsin) in ice-cold 50 mM ammonium bicarbonate. The sample was kept on ice for 45 min. The excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate, and then placed at 37 °C for 15 h. The trypsin digestion was stopped by the addition of 5% formic acid. Distilled de-ionized water (100 µl) was then added, and the tube vortexed gently. The sample was incubated at room temperature for 10 min and then centrifuged for 5 min at 3,000 *xg*. The supernatant was removed and placed in another tube with 5 εl of extraction solution (50% acetonitrile and 5% formic acid). The gel spot was extracted 3 more times with 50 εl of extraction solution and allowed to incubate for 15 min after each extraction. The extracts were combined and placed in a Speed-Vac and evaporated down to 20 εl. The peptide solution was passed through a C18 Spin column; the directions for the Protea C18 kit were followed, allowing for the removal of metal ions from samples (**Protea Biosciences**, Morgantown,WV).

2.2.7. ***MS and MS-MS analysis***

Peptide extracts were sandwiched between 2 layers of matrix. A layer of matrix was dried on the plate (1 εl). Then the sample (1 εl of the above peptide digest) was placed on the matrix and dried. Finally 0.5 εl of matrix was added and dried. The matrix was Ŭ-cyano-4- hydroxy-cinnamic acid (Ŭ-CHCA, 10 mg/ml in 70% acetonitrile/30% 0.1% TFA) (Fluka). For external calibration in the protein mass range, angiotensin I and angiotensin II standards were used. MALDI TOF MS was performed using a Bruker Ultraflex II instrument. MALDI TOF mass spectra were obtained in reflector mode with an acceleration voltage of 25 kV and a pulse ion extraction time of 20 nsec. The mass range for MS was generally between 800- 2700 m/z. The program, Mascot (http://www.matrixscience.com), was utilized in these identifications. MALDI TOF-TOF MS-MS analysis was performed (in the positive ion mode) at the MUSC Proteomics Center using an Applied Biosystems 4700 MALDI MS-MS. MALDI TOF TOF MS-MS was performed on abundant ions observed in MALDI TOF spectra. Peptide sequences were obtained from product ion spectra which are generated for peptide sequences (predicted from the DNA code of sequenced genomes). The experimental peptide spectra are compared to these virtual peptide spectra in the NCBI database and the best matches provided with computer-assistance.

2.3 **Results and discussion**

The strains selected here for study were identified previously as members of the genus *Micrococcus* (using fermentation tests and direct MALDI TOF profiles),but their species identification was less certain. MALDI TOF MS profiling identified all 22 isolates as *Micrococcus luteus*. However, two of the strains (ASO3-C10 and ASO3-C17 were weakly positive for metabolism of glucose, which is inconsistent with the known characteristics of *M. luteus.* The identification scores of these strains were in the range of ñsecure genus identification and probable species identificationòaccording to the Bruker Biotyper software and reference library (Bruker Daltonics, Bremen, Germany), which contains more than 3,200 entries. The assigned scores ranged from 2.00 to 2.299. All of the strains fell within this scoring category. The MALDI TOF MS profiles were compared with data available in a database of reference strains, in order to determine the identification of unknown samples. When the work was performed (2009-2010), the Bruker database contained only one reference of *Micrococcus luteus* ATCC 4698 (Satta et al., 1993).

Physiological tests were performed on gram-positive cocci occurring as quads. Aerobic acid production from glycerol and anaerobic acid production from glucose are two simple tests to distinguish *Micrococcus* from *Staphylococcus* (Kloos, 1974). In the present study, acid production from glycerol and glucose was also assessed. *M. luteus* species are negative for both glycerol and glucose utilization. Two of the 22

environmental isolates produced acid from glucose, and none produced acid from glycerol. As noted above, strains ASO3-C10 and ASO3-C17 were weakly positive for glucose metabolism. These results are consistent with ASO3-C10 and ASO3-C17 being *M. lylae*. However, *Kocuriavarians* also produces acid from glucose but not glycerol (Scherer and Brown, 1974).

In order to confirm and extend the genus and species characterization of the environmental isolates, protein extracts were separated on one dimensional SDS PAGE gels. *M. luteus* ATCC 4698 (and *M. luteus* ATCC 49732) were used as reference strains. Selected gel bands were excised from electrophoretic bands and digested with trypsin. The MALDI TOF spectra were very similar among all environmental isolates (including the 2 strains [ASO3-C10 and ASO3-C17] that displayed aberrant fermentation) and the ATCC reference strains (ATCC 4698), suggesting their identity as *M. luteus* (see Table 1 for list of experimental ions). However, the mass spectrum of ATCC 4698 was the only strain to additionally display the prominent m/z 1623.9 peak. The mass 1623.9 in the other reference was not present in *M. luteus* ATCC 49732. Mass 1623.9 was presumably generated by mutation (K)GVLDV**Q**G**V**EYEIFR(L) (ATCC 4698) to (K)GVLDV**K**G**A**EYEIFR(L) (SK 58). The presence of the additional lysine (SK58) generates an additional Trypsin cleavage site, producing two peptides. These results are consistent with *M. luteus* not being a homogenous species. It has been proposed that *M. luteus* has 3 biovars(Falk and Guering, 1983) based on 16S rRNA sequence comparisons.

An abundant 100 kDa-band was selected for use in MADLI TOF MS identification. Previous work (Rhoden and Miller, 1995) identified a similar sized band in staphylococcal species isolated from indoor air as aconitate hydratase. The identification of a unique aconitate hydratase sequence in *Micrococcus* could provide an additional means for the unequivocal identification of the two genera without the need for physiological tests. Batches of protein extracts from environmental strains were analyzed, together with two ATCC *M. luteus* reference strains (ATCC 4698 [the most frequently used reference standard] and ATCC 49732). *S. warneri* (ATCC 49454), *S. aureus* (ATCC 31240) and *S. hominis* (ATCC 27844) were used as additional control strains.

MALDI TOF MS (for species identification), coupled to MALDI TOF-TOF MS-MS analysis (for protein identification), is a powerful tool for the analysis of tryptic peptides derived from an unknown protein. Signature sequences in proteins can be detected, either through the presence of an amino acid substitution(s) or specific deletions or insertions(Rhoden and Miller, 1995). The ~100 kDa bands detected in all individual environmental isolates and reference strains were digested with trypsin and the mass spectra of tryptic peptides were compared to evaluate similarities and differences. MALDI TOF MS peptide profiles verified that all of the environmental strains had similar mass spectra for this protein, with an abundant and unique peptide being located at 1840.9 m/z (see Fig. .2.1). Using the NCBI database within the MASCOT search engine, commonly used for identifying proteins based on mass spectral data, the ~100

kDa protein was identified as aconitate hydratase. Each environmental isolate was analyzed a minimum of twice, followed by processing the data using MASCOT. In total, 75 MALDI TOF MS analyses were subjected to MASCOT searches on the 22 environmental isolates. Although the spectra were remarkably similar, only 3 of the 75 mass spectra were identified as aconitate hydratase with a MOWSE score, but the value was less than the identification confidence level. Previously, it was noted that MASCOT only identified aconitate hydratase in some of the environmentally isolated staphylococcal strains (Satta et al., 1993), consistent with the results presented here for *Micrococcus*.

Examination of mass spectra of the environmental isolates showed a similar pattern to *Micrococcus luteus* ATCC 49732. All environmental isolates and both *Micrococcus luteus* ATCC reference strains had an abundant ion at 1840.9 m/z (Table 1). They also shared masses 1313.6, 1647.1, 1840.9, 2312.1 but not mass 2013.1 (Table 1). Experimentally, mass 1647.1 was noted to be present in all *M. luteus* strains (including ATCC 4698 and 49732), but only appeared in virtual digests from the sequences derived from the genomes for SK 58 but not from ATCC 4698. None of the *Staphylococcus* strains displayed the mass 1840.9, and all *Staphylococcus* had an abundant m/z of at 2013.1 (Table 2.1) as recorded previously (Fox et al., 2011).

In order to determine whether the tryptic peptides isolated from environmental samples matched those of *Micrococcus luteus*, the aconitate hydratase protein sequence was downloaded from the genomic sequence of *Micrococcus luteus* ATCC

4698 and SK58 (www.ncbi.nlm.nih.gov/bioproject), and a virtual digest of the sequence was performed with Protein Prospector software (prospector.ucsf.edu/prospector). This software takes a known protein sequence and virtually cuts the protein with trypsin. Selecting the type of instrument used in the software, in this case MALDI TOF MS, generates a list of all possible m/z values, which could be present in the mass spectra. Mass spectral peak lists (virtual spectra) obtained with the aid of Protein Prospector were empirically compared with experimental spectra of two *M. luteus* ATCC strains, *S.hominis* ATCC 27844, and 9 environmental samples. The genomic sequences of two strains of *Micrococcus*, *M. luteus* ATCC 4698 (also referred to as NCTC 2665) and *M. luteus* SK58 (2011), are currently present in the GenBank database. These two strains can be separated based on the BLAST distance tree results for both 16S rDNA and aconitate hydratase sequence. In order to verify that the aconitate hydratase sequences generate unique mass spectra from closely related genera, the aconitate hydratase sequences from the most closely related strains, based on the BLAST genomic search, were virtually digested in Protein Prospector and the ions compared. The virtual digest demonstrated that *M. luteus* ATCC 4698 is the only strain that would generate a mass of 1623.8 (see Table 2.2).

Abundant and characteristic masses recorded in MALD TOF MS were selected for further analysis utilizing MALDI TOF-TOF MS-MS. MALDI TOF TOF MS-MS spectra provide an amino acid sequence that allows the identification of the peptide. The MS-MS spectrum provides a breakdown of the peptide with individual masses that

correspond to the most probable amino acid sequence. Based on this amino acid sequence, it is possible to verify the peptide producing the ion for a protein of interest. MS-MS spectra, when processed through databases, such as MASCOT, cannot always identify these proteins. Databases, such as NCBI, may not contain enough examples of species variation, particularly for organisms that are not associated with pathogenesis.

The utilization of tryptic peptides for the identification of uncharacterized environmental bacteria requires caution and empirical interpretation of mass spectral data. Programs, such as Protein Prospector, provide virtual digests linked to given MS-MS parent ions. Comparisons of amino acid sequence ions in virtual fragmentation with experimental MS-MS spectra are invaluable for the comprehensive analysis of the data. As an example, ATCC 4698 on virtual analysis of a 16-mer (mass 1313.6) produced predominantly *y* ions. Of these 16 possible ions, 10 of 16 were observed. For the *b* ions, three were observed in the experimental spectra. The predicted sequence from the experimental MS-MS analysis was IDTPGEAEYYR, which was identical to the sequence generated for virtual MS-MS. Thus, almost complete coverage was observed for this peptide as a component of aconitate hydratase (ATCC 4698 and an environmental isolate).

2.4. **Conclusions**

Characterizing the microbial composition of indoor air is important for a better understanding of disease transmission in a human population. The natural flora found

on human skin contains abundant levels of *Micrococcus* and *Staphylococcus*, which are shed into the air supply of indoor environments. These environmentally-derived organisms have been consistently difficult to classify. Within the *micrococci*, there is at least one well known example of a reference strain being mis-categorized (Baldellon and Megraud, 1985), which led to an inaccurate reference standard for Food and Drug Administration (FDA)-mandated testing. In environmental indoor air samples, it was found previously that over two thirds of the environmental isolates were *micrococci* (BrukerBiotyping), creating a need for greater accuracy in characterization (Satta et al., 1993). Characterization of fermentation characteristics and MALDI TOF profiles allows that categorization of environmental *Micrococcus* isolates to the genus level. Tryptic peptide analysis (e.g., aconitate hydratase) by MALDI TOF MS profiles provides other criteria for a confirmation of the species identity for environmental isolates. MALDI TOF-TOF MS-MS can be used to identify proteins from these peptides. It has been reported that *M*. *luteus* may include more than one biovar, based on 16S rDNA sequence data (Falk and Guering, 1983). Comparisons of mass spectral data of tryptic peptides also show a clear difference between the profiles of the two deposited *M. luteus* (ATCC4698 and 49732), although the environmental strains were remarkably similar. Further investigations are warranted to determine whether these two ATCC strains and possibly others (displaying an amino acid substitution in aconitate hydratase) are distinct biovars or species within the genus *Micrococcus*.

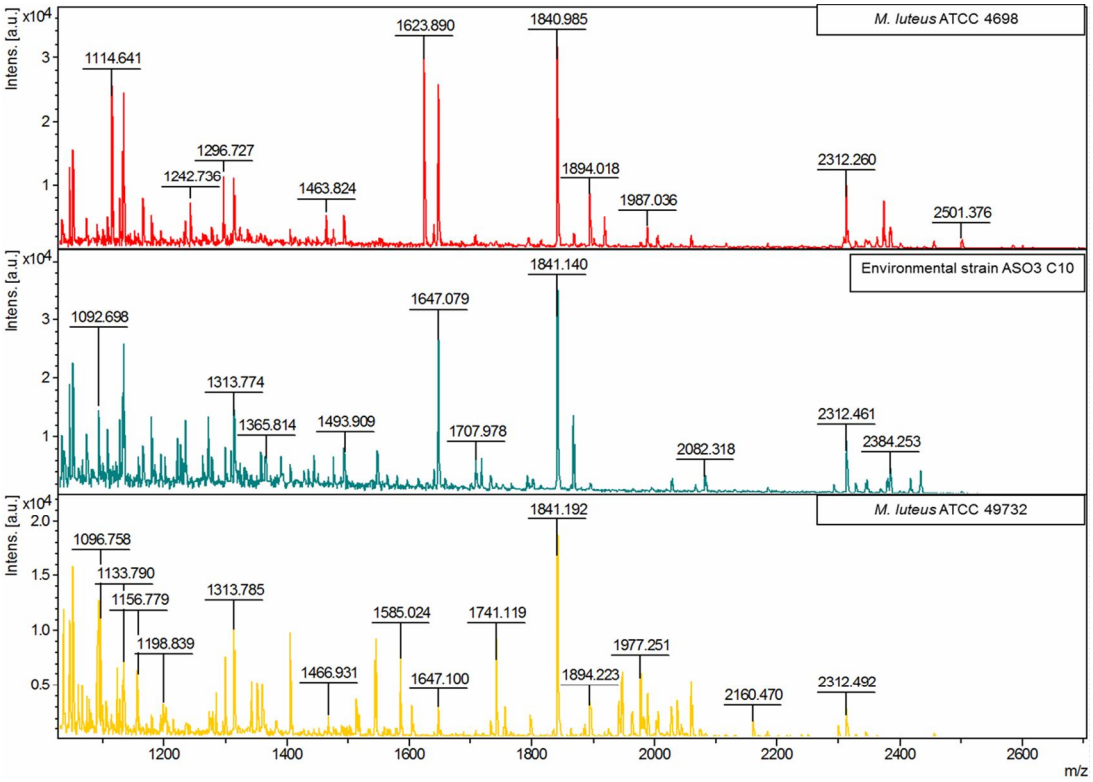


Figure 2.1 Mass spectra of an ATCC strain (*M. luteus* 4698) and two representative environmental isolates. The major peaks were found in all strains. The prominent 1623.9 m/z peak was found only in the ATCC 4698 strain of *M. luteus*.

Table 2.1 MS ions present in strains

	969.6	1214.5	1313.6	1405.7	1623.8	1840.9	2013.1
M.L. ATCC 4698	(+)	(+)	(+)	(+)	(+)	(+)	ND
M.L. ATCC 49732	(+)	(+)	(+)	(+)	ND	(+)	ND
ASO3 c5 upper	(+)	(+)	(+)	(+)	ND	(+)	ND
ASO3 c6 upper	(+)	(+)	(+)	(+)	ND	(+)	ND
ASO3 c10 upper	(+)	(+)	(+)	(+)	ND	(+)	ND
Uno c10Y	(+)	(+)	(+)	(+)	ND	(+)	ND
ASO3 c 45	(+)	(+)	(+)	ND	ND	(+)	ND
ASO3 c31	ND	(+)	(+)	(+)	ND	(+)	ND
ASO3 c 17	ND	(+)	(+)	(+)	ND	(+)	ND
Micro Sp.	(+)	ND	ND	(+)	ND	(+)	ND
ASO3 c 19	(+)	ND	ND	(+)	ND	(+)	ND
ASO15 c 31	(+)	ND	ND	(+)	ND	(+)	ND
ASO3 c 15	(+)	(+)	ND	(+)	ND	(+)	ND
Uno 2 w	(+)	ND	(+)	(+)	ND	(+)	ND
S. hominis	ND	ND	ND	ND	ND	ND	(+)
S. warneri	ND	ND	ND	ND	ND	ND	(+)
S. aureus	ND	ND	ND	ND	ND	ND	(+)

CHAPTER 3

Identification of Staphylococcal species based on variations in protein sequence (Tandem Mass Spectrometry) and DNA sequence (Microarray)

Jennifer M. Kooken, Karen F. Fox, Alvin Fox*

DEPARTMENT OF PATHOLOGY, MICROBIOLOGY AND IMMUNOLOGY, SCHOOL OF MEDICINE, UNIVERSITY OF SOUTH CAROLINA, COLUMBIA, SC 29208, USA

(Manuscript in Progress)

* CORRESPONDING AUTHOR. TEL. 803 216 3409; FAX 803 216 3428.

E-MAIL: ALVIN.FOX@USCMED.SC.EDU.

3.0 **Abstract**

This report is among the first and most definitive studies using sequence variation in newly discovered protein markers for bacterial speciation. Variation, at the DNA level, in the *sodA* gene was used for comparison; this is an established technique used for staphylococcal speciation. A total of 64 strains were analyzed including 12 reference strains (representing 10 CoNS species and *S. aureus*, see Table 1). MALDI TOF MS or LC ESI MS-MS were used for peptide analysis of proteins isolated from gel bands. Visual comparison of experimental spectra of unknowns versus spectra of peptides derived from reference strains allowed bacterial identification after MALDI TOF MS analysis. After LC-MS-MS analysis of gel bands bacterial speciation was performed in an automated fashion comparing experimental spectra versus virtual spectra using the software X!Tandem. In the final series of experiments LC-MS-MS was performed on whole proteomes and data analysis also employing X!tandem. Aconitate hydratase and oxoglutarate dehydrogenase served as marker proteins on focused analysis after gel separation. Alternatively on full proteomics analysis elongation factor tu provided the highest confidence in staphylococcal speciation.

3.1 **Introduction**

Staphylococcal infections, including life threatening bacteremia and endocarditis are among the most common hospital acquired infectious diseases. The current work is concerned with the demanding task of providing correct species identification of coagulase negative staphylococci species of human origin. *Staphylococcus aureus* (coagulase positive) and *Staphylococcus epidermidis* (coagulase negative, CoNS) are the two staphylococcal species most frequently isolated in human infections. Unlike other CoNS, it has been observed that *S. lugdunensis* infections resemble those mediated by *S. aureus* in terms of aggressiveness and severity of the infection, increasing the significance of accurately identifying *S. lugdunensis* (Herchline and Ayers, 1991; Mateo et al., 2005; Noguchi et al., 2010; Pereira et al., 2010; Shah et al., 2010). *S. saprophyticus* is generally only associated with urinary tract infections. *S. hominis*, *S. warneri, S. capitis*, *S. haemolyticus*, *S. lugdenensis*, *S. simulans and S. cohnii*constitutemost of the remaining CoNS species isolated from man (Center et al., 2003; Frank et al., 2004; Jukes et al., 2010; Ohara-Nemoto et al., 2008; Sivadon et al., 2005). Batteries of biochemical tests are still the primary means of staphylococcal species identification (Marrie et al., 1982). However, it is also well known that physiological tests do not provide accurate speciation of CoNS.

Indeed *S. lugdunensis* and *S. haemolyticus* are often lumped together as are *S. hominis*/*S. simulans* and *S. capitis*/*S. epidermidis* (Skow et al., 2005). Furthermore, it is repeatedly stated that *S. hominis* is misidentified using commercial phenotypic tests

being commonly confused with *S. warneri or S. epidermidis* (Fujita et al., 2005; Gilad and Schwartz, 2007; Spanu et al., 2003; Yugueros et al., 2000).

In an attempt to address the issue of misidentification of *staphylococcal* species, variation in the sequence of 16S rRNA has been widely employed. Unfortunately, it is now recognized that 16S rRNA is highly conserved among staphylococcal species (above 98% similarity) and is not usually helpful in their discrimination. As a consequence sequencing of a variety of other more variable genes has been employed as an alternative to 16S rRNA including *rpoB*, *hsp60*, *sodA*, and *dnaJ* genes respectively displaying 86%, 82%, 81%, and 77% similarity (Goh et al., 1997; Hirotaki et al., 2011; Kwok et al., 1999). 16S-23S rDNA intergenic spacer (ISR) PCR has also been utilized. However among *staphylococci*, multiple ISR copies are expressed and ISR patterns require expert evaluation (Fujita et al., 2005; Mendoza et al., 1998).

Variation in the sequence of the *sodA* gene has been the most widely considered for use in the clinical microbiology laboratory and both real time PCR and microarray techniques have been employed (Giammarinaro et al., 2005; Iwase et al., 2007; Poyart et al., 2001; Sivadon et al., 2005). However, to this point validation of the utility of sodA variation with another independent approach has remained elusive. For example, in a recent report it was stated that MALDI TOF (matrix assisted laser desorption/ionization time-of -flight) mass spectrometry (MS) mass profiling correlated best with 16S rRNA but not *sodA* sequence (Dubois et al., 2010). Indeed other reports have suggested that direct MALDI TOF MS does indeed discriminate many organisms isolated in clinical microbiology laboratories. Unfortunately *staphylococci* are particularly difficult to identify at the species level (Carbonnelle et al., 2012); these workers suggested a more specific

method identifying variations in protein sequence (distinct from direct MALDI-TOF MS) has the potential for greater accuracy (Fox et al., 2010, 2011).

In our previous report, a limited battery of strains representing 6 species (*S. capitis, S. chromogenesS. cohnii. S. hominis S. saprophyticus* and *S. warneri)* derived from human skin and nares where characterized using variation in protein sequence of the enzyme aconitate hydratase). Accordingly in the current report, a much larger group of staphylococcal strains representing 11 *staphylococcal* species strains; *S. aureus S. capitis*, *S. cohnii*, *S. epidermidis, S. haemolyticus*, *S. hominis, S. lugdunensis,S. saprophyticus, S. simulans,* and *S. warneri*were studied . Determination of variation in peptide sequence of specific marker proteins included the use of liquid chromatography-electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS). Previously we employed sequence variation in an enzyme (aconitate hydratase) for staphylococcal speciation (Fox et al., 2011). However here the utility of other proteins (including oxoglutarate dehydrogenase and elongation factor Tu) were additionally used as species markers. Sequence variation of the *soda* gene (assessed using microarrays) was compared for mutual validation.

3.2 **Materials and Methods**

3.2.1 ***Strains characterized***

A total of 64 strains were characterized which included 12 reference strains (11 different staphylococcal species): *S. aureus* (ATCC 12598, Cowan 1)*, S. capitis* (ATCC 27841 and ATCC 35661), *S. cohnii*(ATCC 29972)*, S. epidermidis* (ATCC 12228)*, S. haemolyticus*(ATCC 29970), *S. hominis* (ATCC 27844)*, S. lugdunensis* ATCC 49576*, S. saprophyticus* ATCC 15305)*, S. simulans* (ATCC 27851)and *S. warneri*(ATCC 49454). Other strains included 25 clinical samples of human origin, 19 strains isolated by air sampling from occupied school rooms (associated with shed human skin, Fox et al., 2010) and eight strains of veterinary origin (see Table 3.1).

3.2.2 ***Culture conditions and confirmation as staphylococci***

Bacteria were grown on nutrient agar plates at 37 °C for 24-48 h. Staphylococci were differentiated from micrococci by glucose fermentation and genus identity confirmed by mass spectrometry (Fox et al., 2011, Kooken et al. 2012).

3.2.3 ***Protein separation, tryptic digestion and MS analysis***

Bacteria were harvested (after growing as confluent lawns) from plates using 0.1 M NaCl, 50 mM Tris HCl, 0.5 mM phenylmethylsulfonyl fluoride [PMSF] and placed in a

FastPrep®-24 bead beater (MP Biomedicals, Solon, OH) for 6 ms X 30 sec with 5 min on ice between each cycle for a total of 6 cycles to release proteins. The samples were then centrifuged at 4 °C for 1 h at 10,000 *g*. The supernatants were removed and placed at -70 °C for two freeze-thaw cycles to eliminate DNA.

Prior to electrophoresis samples were vortexed, and 20 εl of each supernatant was added to 20 εl of 2x loading buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M tris-HCl), and boiled at 100 °C for 5 min then centrifuged for 5 s. Samples were subjected to electrophoresis in 5% Criterion Tris-HCl gels (Bio-Rad) until the bromophenol blue was at the gel bottom. The running buffer consisted of 192 mM glycine, 25 mM Tris and 0.1% SDS. Gels were washed 3 times in water for 5 min each and stained with Gel-Code Blue Stain reagent (Pierce, Rockford, IL) for 1 h.

Bands of interest were excised from Coomassie blue-stained gels and de-stained with 25 mM ammonium bicarbonate in 50% acetonitrile/50% water. Each gel was dehydrated by addition of acetonitrile for 5 min. The acetonitrile was removed and the gel rehydrated with 100 mM ammonium bicarbonate (10 min). Then, the gel was dehydrated with acetonitrile (5 min), the acetonitrile removed and the sample dried for 2-3 min. The sample was rehydrated with 10 mM dithiothreitol (DTT) and incubated for 30 min at room temp (22-24 °C), then alkylated with 50 mM iodoacetamide for 30 min at room temp, and then washed in 100 mM ammonium bicarbonate. Once the iodoacetamide had been removed, the sample went through 3 further rounds of

dehydration-rehydration (100% acetonitrile then 100 mM ammonium bicarbonate in water). After the final dehydration, the sample was dried. The gel band was rehydrated (with a 12.5 ng/ml solution of trypsin) in ice-cold 50 mM ammonium bicarbonate. The sample was kept on ice for 45 min. The excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate, and then placed at 37 °C for 15 h. The trypsin digestion was stopped by the addition of 5% formic acid. Water (100 µl) was then added, and the tube vortexed gently. The sample was incubated at room temp for 10 min and then centrifuged for 5 min at 3,000 *g*. The supernatant was removed and placed in another tube with 5 εl of extraction solution (50% acetonitrile and 5% formic acid). The gel spot was extracted 3 more times with 50 εl of extraction solution and allowed to incubate for 15 min after each extraction. The extracts were combined and and evaporated down to 20 εl.

For MALDI TOF MS analysis, the peptide solution was passed through a C18 Spin column using a Protea C18 kit for removal of metal ions from samples (Protea Biosciences, Morgantown, WV). Peptide extracts were sandwiched between 2 layers of matrix. A layer of matrix (Ŭ-cyano-4- hydroxy-cinnamic acid (Ŭ-CHCA, 10 mg/ml in 70% acetonitrile/30% 0.1% TFA) was dried on the plate (1 εl). Then the sample (1 εl of the above peptide digest) was placed on the matrix and dried. Finally 0.5 εl of matrix was added and dried. MALDI TOF MS was performed using a Bruker Ultraflex II instrument. MALDI TOF mass spectra were obtained in reflector mode with an acceleration voltage of 25 kV and a pulse ion extraction time of 20 nsec. The mass range for MS was

generally between 800- 2700 m/z. Species were identified by visual comparison of unknowns to spectra of ATCC reference strains.

For LC-ESI MS/MS trypsin-digested peptides from gel bands (as described above) were also processed using solid phase extraction (SPE). SPE was performed with a vacuum manifold using Strata C-18 T solid phase extraction columns (Phenomenex, Torrence, CA). Briefly, 1 ml of 100% methanol was added to activate the resin, followed by a conditioning step of 1ml 0.1% TFA water, then addition of the samples. The samples were washed with 5% acetonitrile in 0.1% TFA water, and finally elution of the samples with 80% acetonitrile in 0.1% TFA in water. Samples were dried down to near completeness (5-10 µl remaining) with a Thermo speed vac. 25µl of 0.5% formic acid was added to each sample.

In the last set of experiments, bacterial supernatants prepared as described above were subjected directly to tryptic digestion and C18 column clean-up prior to LC-MS-MS analysis. 50 µl of supernatant (prepared as above), 50 µl 8M urea, 1 µl of ƀ-mercaptoethanol, 24 µl of water, and 25ul of 200 mM Tris-HCl pH 8.0 were added tubes mixed and incubated at 60°C for one hour in a Thermomixer (ThermoFisher Scientific, Billerica, MA) shaking at 300RPM. 800 µl of 50mM ammonium bicarbonate was added to each tube to reduce the urea concentration to below 1M. 2 µl of trypsin gold at a concentration of 2 µg/µl was added to each tube and briefly vortexed to mix. The samples were then incubated at 37°C for 15 hours in a Thermomixer shaking at 300

RPM. The samples were then subjected to clean-up with the Strata C-18 T solid phase extraction columns as described above.

For LC-MS-MS analysis, samples were processed using an auto sampler and an LC- ESI LTQ Orbitrap LX (Thermofisher) equipped with a fused silica capillary column (40 cm length, I.D. 0.15 mm) packed with Jupiter 5 εm C-18 resin with a flow rate of 2 εl/min using a 2 solvent flow system (solvent A: 5% acetonitrile, 0.1% formic acid in water; solvent B 90% acetonitrile; 0.1% formic acid in water). Electrospray conditions were 3.5 kV spray voltage, 200 ^{0}C source temperature and 200 V ion tube transfer voltage.

Database searches using ESI-MS-MS data were performed using the freeware program X!Tandem (www.thegpm.org/tandem). The data base was modified by downloading the genomes from all eleven species representing the 64 strains in this study. The software analysis is described in greater detail in the next chapter.

*3.2.4**Microarray probes for the sodA gene***

The sodA gene was amplified and labeled with Cy3 by PCR from DNA isolated from staphylococcal reference strains, human skin and veterinary samples. Reactions contained in a total volume of 40 µl; DNA (160 ng), ThermoPol Buffer (1X), Taq DNA Polymerase (1 U , New England BioLabs) dATP (1.0 mM), dTTP (1.0 mM), dGTP (1.0 mM), and dCTP (0.1 mM, Invitrogen), Cy3-dCTP (0.9 mM GE Healthcare), D1 and D2

primer mixes (10 µM) and 5% DMSO The PCR protocol was 3 min at 95^{0}C, 50 cycles

of 30 sec at 95^{0}C, 1 min at 37^{0}C and 45 sec at 72^{0}C, and 6 min at 72^{0}C. PCR products were purified using the GeneElute PCR Clean Up Kit (Sigma-Aldrich) according to the manufacturerôs recommendations). The protocol was derived from Poyart et al. 2001.

Spotting of microarrays, hybridization, and scanning: DNA probes for *S. aureus* and CoNS species were described in Giammarinaro et al., 2005. Probes were diluted to a final DNA concentration of 30 µM in 3X SSC buffer and printed onto UltraGAPS coated slides (Corning Inc., Cat. No. 40015) using the non-contact piezoelectric Biochip Arrayer (PerkinElmer Inc.). Eight arrays were printed per slide with each array having all 13 DNA probes printed in quintuplicate. After printing the slides were pre-hybridized using the Pronto Universal Microarray Hybridization Kit (Corning Inc) according to the manufacturerôs recommendations. Amplified and Cy3 labeled PCR product of the sodA gene were hybridized to the DNA probes on the slides using the Agilent Gene Expression Hybridization kit. Briefly, 20 ng of Cy3 labeled DNA were mixed with 10X blocking buffer and denatured for 2 min at 95^{0}C. Hybridization buffer was then added

and the hybridization mixes were incubated with the arrays for 17 h at 58^{0}C with mixing.

Slides were washed using the ñGene expression wash buffer kitò (Agilent Technologies) and dried by immersing and slowly removing slides from 100% acetonitrile. Slides were scanned using a high resolution DNA microarray scanner (Agilent Technologies, Inc.) at 5µm resolution

3.3 **Results**

This represents one of the first studies using sequence variation in newly discovered protein markers for bacterial speciation. Variation, at the DNA level, in the *sodA* gene was used for comparison; this is an established technique used for staphylococcal speciation. As noted above a total of 64 strains were analyzed including 12 reference strains (representing 10 CoNS species and *S. aureus*, see Table 3.1). MALDI TOF MS or LC ESI MS-MS were used for peptide analysis of proteins isolated from gel bands. Visual comparison of experimental spectra of unknowns versus spectra of peptides derived from reference strains allowed bacterial identification after MALDI TOF MS analysis. After LC-MS-MS analysis of gel bands bacterial speciation was performed in an automated fashion comparing experimental spectra versus virtual spectra using the software X!Tandem. In the final series of experiments LC-MS-MS was performed on whole proteomes and data analysis also employing X!tandem. *S. simulans* was found to be the dominant species of *staphylococcus* isolated from clinical isolates; indeed 16 of the 25 human isolates were found to be *S. simulans*. Another species *S. lugdenensis* (5 strains) constituted most of the remaining strains. These two species were not commonly found among human skin isolates or veterinary strains. Another commonly isolated species that was still difficult to identify before this study was initiated is *S. lugdenensis.* In comparison to *S. simulans* the banding pattern on SDS-PAGE for *S. lugdenensis* exhibited a triplet of bands above the 100 kD marker whereas for *S. simulans* a doublet was observed. As noted above, the band migrating

above the 100 kD marker in other *Staphylococcus* species was previously identified as aconitate hydratase and therefore the corresponding band for *S. lugdenensis* was selected for analysis. The *S. lugdenensis* band did not display the characteristic peptide (mass 2013) universally observed in staphylococci. Strains CNS 5, 6, 8, 12, and 17 were identified as *S. lugdenensis* utilizing the oxoglutarate dehydrogenase marker protein, Among the dominant masses observed in MALDI TOF MS is spectra is mass 2340.; this mass was also observed in virtual spectra for oxoglutarate dehydrogenase generated with the program Protein Prospector (http:prospector.ucsf.edu/prospector). LC-MS-MS observations confirmed the dominant protein for *S. lugdenensis* as oxoglutarate dehydrogenase (see Table 3.2).

In summary on analysis of 100 kDa gel bands, the aconitate hydratase marker for staphylococci previously described (Fox et al., 2011; Kooken et al 2012) was prominent for most CoNS species including *S. simulans* characterized for the first time in the present study. However oxoglutarate dehydrogenase was the dominant band for *S. lugdenensis* . In contrast for whole proteomes, peptides derived from aconitate hydratase and oxoglutarate dehydrogenase were buried in the complex mixture of peptides derived from other proteins. For most strains, alternatively, the protein found to provide the top hit for bacterial speciation was elongation factor tu. Less commonly two other proteins, enolase and ATP-dependent Clp protease, provided bacterial identification with the highest confidence (see Table 3.3).

As noted above, for the clinical samples, most strains were found to be *S. simulans* (16/25 CoNS) or *S. lugdenensis* (5/25). Using sodA microarray, for 6 strains (including the reference strain, ATCC49576) the strongest signal on probe binding

corresponded to *S. lugdenensis*. For CNS 17 only binding to the *S. lugdenensis* probe was observed. In 4 cases (CNS 5, 6, 8, 12) binding to the *S. hominis* probewasobserved asa secondary signal. In only one case (CNS10) was a secondary signal observed corresponding to *S. simulans*. CNS10 was found to be *S. simulans* on mass spectrometry analysis.

Five strains were not identified on microarray analysis: CNS11, ASO15-C40Y, AS015-C28, AS015-C106, MUS 5951. CNS-11, ASOC15-C28, ASO15 C106 were identified by proteomics as *S. haemolyticus.* In an earlier study we identified ASOC15-C28, ASO-C15 ASO-C106 as *S. chromogenes* based on similar aconitate hydratase mass spectra. However, a more limited species data base was used employing only 6 species. *S. haemolyticus* was not among the reference strains studied. *S. haemolyticus* is commonly isolated from human samples but *S. chromogenes* is generally only found in veterinary samples. The *S. haemolyticus* reference strain gave a weak signal compared to the other 11 species analyzed on microarray analysis which may explain the negative results for the more divergent unknowns. The probe for *S. haemolyticus* is among the short ones designed for staphylococcal identification using the sodA gene which may explain the weak signal. Varying hybridization conditions, including lowering the binding temperature still did not positive result on array analysis. Three other strains of *S. chromogenes* including the reference strain and 2 veterinary isolates were correctly identified on microarray analysis. ASO15-C40Y C40Y was identified with the lowest score by MS-MS as *S. simulans*. Thus it may represent an uncommon species not included in our study. MUS591 was identified using proteomics as S. warneri. There were only 3 strains where there are disagreements between MS

and arrays - CNS 10 (*S. simulans* [MS] and *S. lugdenensis* [array]), AS03 59W (*S. lugdenensis* [MS] and *S. warneri* [array]) and Cow924RR (*S. capitis* [MS] and *S. saprophyticus* [array]). According the sodA gene for the eight strains (CNS-11, ASOC15-C28, ASO15-C106, C40Y, MUS5951, CNS10 and AS03 59W) is currently being sequenced to provide further information on their identity.

3.4**Discussion**

This study established that variation in protein sequence assessed by sequencing trypsin-released peptides using and tandem mass spectrometry correlated for species identification of staphylococcal isolates of human origin. It is stressed that the well-known õdirectò MALDI TOF MS approach provides a whole cell protein pattern, without identification of any individual protein or their sequence variation. Direct MALDI TOF MS was not employed in the current study, since it has proven difficult to discriminate staphylococcal species with this technique. The MS-MS results correlated well with 66 sodA sequence determined using microarrays).

MALDI TOF MS, for bacterial speciation, was originally described by several groups almost 20 years ago and has only come to fruition as a clinical microbiology tool in the past few years. Unfortunately MALDI TOF MS does not always have the desired specificity for the complex issue of staphylococcal speciation although MALDI TOF MS profiling is useful in identifying many other bacteria. The developmental work described here was not designed to provide tests suitable for use in the clinical microbiology laboratory at the current time.

However the current study is a significant step forward in the process of fulfilling the hitherto elusive goal of reliable speciation of CoNS derived from clinical specimens. At this time, we do not recommend microarrays nor mass spectrometry for routine use in the clinical laboratory; both techniques remain too complicated to satisfy the need for simple and high through-put analysis. Microarrays allow all staphylococcal species to be differentiated simultaneously with their own sequence specific probe (based on previously discovered sequence variation). Tandem mass spectrometry, as demonstrated here, is flexible and allows new protein marker sequences to be identified as the need arises. Technology development (including simplification of sample preparation and improvements in bioinformatics for data handling for both microarrays and proteomics are both active areas of research and it may be expected to see applications in staphylococcal speciation (and also for other genera) designed for the clinical microbiology laboratory in the not too distant future.

The current study was also not focused on the role of particular CoNS species associated with specific human infections or distinctions from the normal flora which can often contaminate clinical samples. There have been numerous studies carried out by others addressing these important issues. However the only commercial tests widely available have been batteries of biochemical profiles. Indeed there may be a need for additional future studies using more specific-methods, as described here, to allow focus on specific CoNS species (e.g. *S. lugdenensis*) now that they can be differentiated from other common human isolates (e.g. *S. simulans*) where there is less evidence for importance in human infection.

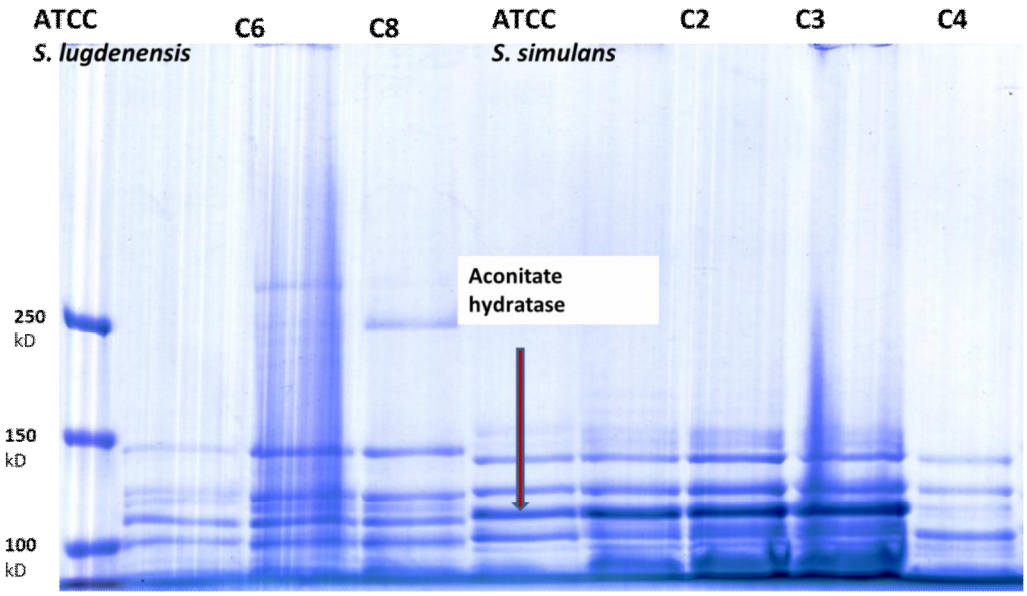


Figure 3.1 ID gel separation of protein mixtures extracted from *S. simulans* and *S. lugdenensis* strains.

Table 3.1 Identification of staphylococcal species using mass spectrometry (protein sequence variation) and microarray (69 sodA DNA sequence variation)

Strain	Origin	Identification: mass spectrometry	Identification: microarray
CNS1	Clinical	*S. simulans*	*S. simulans*
CNS2	Clinical	*S. simulans*	*S. simulans*
CNS3	Clinical	*S. simulans*	*S. simulans*
CNS4	Clinical	*S. simulans*	*S. simulans*
CNS5	Clinical	*S. lugdenensis*	*S. lugdenensis*
CNS6	Clinical	*S. lugdenensis*	*S. lugdenensis*
CNS7	Clinical	*S. simulans*	*S. simulans*
CNS8	Clinical	*S. lugdenensis*	*S. lugdenensis*
CNS9	Clinical	*S. warneri*	*S. warneri*
CNS10	Clinical	*S. simulans*	*S. lugdenensis*
CNS11	Clinical	*S. haemolyticus*	
CNS12	Clinical	*S. lugdenensis*	*S. lugdenensis*
CNS13	Clinical	*S. simulans*	*S. simulans*
CNS14	Clinical	*S. simulans*	*S. simulans*
CNS15	Clinical	*S. simulans*	*S. simulans*
CNS16	Clinical	*S. simulans*	*S. simulans*
CNS17	Clinical	*S. lugdenensis*	*S. lugdenensis*
CNS18	Clinical	*S. aureus*	*S. aureus*
CNS19	Clinical	*S. simulans*	*S. simulans*
CNS20	Clinical	*S. aureus*	*S. aureus*
CNS21	Clinical	*S. simulans*	*S. simulans*
CNS22	Clinical	*S. simulans*	*S. simulans*
CNS23	Clinical	*S. simulans*	*S. simulans*
CNS24	Clinical	*S. simulans*	*S. simulans*
CNS25	Clinical	*S. simulans*	*S. simulans*
ASO3-C19	Human skin	*S. hominis*	*S. hominis*
ASO3-C22	Human skin	*S. saprophyticus*	*S. saprophyticus*
ASO3-C73	Human skin	*S. hominis*	*S. hominis*
ASO3-C77	Human skin	*S. warneri*	*S. warneri*
ASO15-C40Y	Human skin	*S. simulans*	
ASUNO15-C6y	Human skin	*S. aureus*	*S. aureus*
ASO1-C8	Human skin	*S. aureus*	*S. aureus*
ASUNO2/15-C6y	Human skin	*S. aureus*	*S. aureus*
ASUNO2/15-C11y	Human skin	*S. warneri*	*S. warneri*
AS015-C28	Human skin	*S. haemolyticus*	
AS015-C106	Human skin	*S. haemolyticus*	
ASO2 C44	Human skin	*S. simulans*	*S. simulans*
ASC02 C53	Human skin	*S. epidermidis*	*S. epidermidis*
ASO-15 C64	Human skin	*S. aureus*	*S. aureus*
ASO15-C84	Human skin	*S. aureus*	*S. aureus*
AS02 C58W	Human skin	*S. epidermidis*	*S. epidermidis*
AS02 C63Y	Human skin	*S. epidermidis*	*S. epidermidis*
AS03 C6	Human skin	*S. aureus*	*S. aureus*
ASO2-C21	Human skin	*S. warneri*	*S. warneri*
AS03 59W	Human skin	*S. lugdenensis*	*S. warneri*
Cow990RR	Veterinary	*S. simulans*	*S. simulans*
09-200-150	Veterinary	*S. warneri*	*S. warneri*
MUS 5949	Veterinary	*S. chromogenes*	*S. chromogenes*
09-304-034	Veterinary	*S. saprophyticus*	*S. saprophyticus*
Cow 924RR	Veterinary	*S. capitis*	*S. saprophyticus*
Cow970RR	Veterinary	*S. saprophyticus*	*S. saprophyticus*
MUS 5951	Veterinary	*S. warneri*	

Table 3.2 Identification of the most protein the highest confidence score isolated from 100 kDa gel bands using liquid chromatography-tandem mass spectrometry

Strain	**Score**	**Protein**
S. capitis ATCC 27841	-440.7	Aconitate hydratase
*S. haemolyticus*ATCC 29970	-380.2	Aconitate hydratase
S. lugdenensis ATCC 49576	-428.3	Oxoglutarate dehydrogenase
S. saprophyticus ATCC 15305	-288.9	Aconitate hydratase
S. simulans ATCC 27851	-423.1	Oxoglutarate dehydrogenase
CNS 9	-277.9	Aconitate hydratase
CNS 10	-476.1	Aconitate hydratase
CNS 13	-327.1	DNA polymerase
CNS 17	-248.1	Oxoglutarate dehydrogenase
AS02 c58W	-101.0	Aconitate hydratase
CNS11 100K	-222.3	Aconitate hydratase

Table 3.3 Identification of the protein with the highest confidence score in whole staphylococcal proteomes using liquid chromatography-tandem mass spectrometry

Strain	**Score**	**Protein**
S. aureus ATCC 12598	-276.2	Elongation factor Tu
S. capitis ATCC 27841	-357.9	Elongation factor Tu
S. cohnii ATCC 29972	-255.7	Elongation factor tu
S. chromogenes ATCC 43764	-208.8	Elongation factor tu
S. epidermidis ATCC 12228	-441.7	Elongation factor Tu
S. lugdenensis ATCC49576	-372.3	Elongation factor Tu
S. simulans ATCC 27851	-372.3	Elongation factor Tu
CNS 1	-219.6	Elongation factor Tu
CNS5	-371.4	Elongation factor Tu
CNS6	-468.1	Elongation factor Tu
CNS7	-489.8	ATP-dependent Clp protease
CNS8	-450.4	Elongation factor Tu
CNS10	-525.8	Elongation factor Tu
CNS18	-227.3	Elongation factor Tu
CNS20	-299.4	Elongation factor Tu
aso2c21	-459.5	Elongation factor Tu
aso15c28	-338.2	Elongation factor Tu
aso15c40y	-102.5	Elongation factor Tu
aso15c106	-366	Elongation factor Tu
aso1c8	-243.1	Elongation factor Tu
aso2c53	-353.5	Elongation factor Tu
aso2c63	-347.2	Elongation factor Tu
aso3c6	-230.2	Enolase
aso3c59w	-277.9	Elongation factor Tu
ASO3 C22	-411.5	Elongation factor Tu
ASO2 C44	463.8	ATP-dependent Clp protease
aso15c64	-226.8	Enolase
aso15c84	-217.6	Enolase
asuno2-15c6y	-185.9	Elongation factor Tu
asuno15c6y	-270.2	Enolase
cow924RR	-169.2	Elongation factor Tu
cow970RR	-400.8	Elongation factor Tu
mus5949	-296.1	Elongation factor tu
mus5951	-150.4	Elongation factor Tu
09-304-034	-315.7	Elongation factor Tu
900-200-150	-344.9	Elongation factor Tu
990RR	-245.6	Elongation factor Tu

CHAPTER 4

Liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI MS/MS) for bacterial speciation using high-through-put proteomics

Jennifer M. Kooken, Karen F. Fox, Alvin Fox*

Department of Pathology, Microbiology and Immunology, School of Medicine, University of South Carolina, Columbia, SC 29208, USA

(Manuscript in Progress)

* Corresponding author. Tel. 803 216 3409; fax 803 216 3428.

E-mail: alvin.fox@uscmed.sc.edu.

4.0 **Abstract**

A collection of coagulase negative staphylococcal strains (CoNS) originating from normal human skin or clinical samples (human and veterinary origin) were specieated in this study. Digests of proteins released from whole cells were converted to tryptic peptides for the determination of species. This technique eliminated time-consuming steps involving 1D gel separation and protein extraction from gel slices previously employed. Liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS, Thermo Fisher Orbitrap) was employed for peptide analysis. Multiple proteins were identified based on peptide sequences and used to identify species of *Staphylococcus*. However the top hit (elongation factor tu or enolase) alone were found to provide accurate speciation. Markers previously identified from gel slices (aconitate hydratase and oxoglutarate dehydrogenase) were poor markers in whole cell digests. Bioinformatics was performed employing the software X!Tandem which uses sequenced genomes to generate a virtual peptide database for comparison to experimental data. The search database was modified to utilize the genomes of the 11 *Staphylococcus* species most commonly isolated from man. The methodological approach described here provided a simple yet elegant way of identification of staphylococci. However, perhaps more importantly the methodology is applicable universally for identification of any bacterial species.

4.1 **Introduction**

Proteomics is primarily utilized in identifying proteins expressed by an organism under particular growth conditions, not for chemotaxonomic characterization. MALDI-TOF MS profiling has become a routine technique for identification of bacteria particularly with relevance to clinical microbiology MALDI-TOF MS is used for rapid determination of a mass pattern of proteins for bacterial characterization; these proteins are generally not identified (Intellicato and Fox, 2013). Alternatively, there have been a handful of reports identifying bacteria more accurately by comparing experimental spectra of tryptic peptides using either LC-MS-MS and custom bioinformatics software (Jabbour et al., 2010) or MALDI TOF MS and MS-MS of isolated gel bands (Fox et al., 2011).

Microbiological testing in clinical settings is still usually based on culture followed by Gram stain morphology and biochemical characteristics (Baker, 1984; Gupta, 1998). Physiological traits additionally evaluate various properties including the ability to ferment different substrates (including sugars) and enzymatic activity, e.g. hemolysis or coagulase activity (Noble, 1984). Many of these tests are routinely used for accurate identification of many pathogenic or opportunistic species but for less studied species (e.g. coagulase negative staphylococci, CoNS) the results are often inconclusive. Additionally technical problems such as incomplete Gram staining, or partial fermentation of sugars can lead to erroneous species identification (De Paulis et al., 2003).

Numerous variants of the polymerase chain reaction (PCR) and/or restriction enzymes are commonly used in more advanced reference laboratories in species identification. However in developing new genetic markers two conserved genetic regions (surrounding a variable region) are required for primers to provide initiation for amplification and for the variable regions to provide the information for discriminating closely related species. When limited information is available, selection of a single gene for assessing sequence variation can be somewhat arbitrary. Whole genome sequencing is still a technically demanding and expensive alternative. Accordingly many workers use genes that are present universally in bacteria, most commonly16S rDNA. However, it has become clear over the past few years that 16S rDNA sequences are too conserved amongst many closely related species and only genus-level identification is provided (Brodie et al., 2007).

Alternatively our group has employed members of the genus *Staphylococcus* as a model system to improve accuracy for speciation of difficult-to-identify species using protein markers (Fox et al., 2010; Fox et al., 2011, Kooken, Fox and Fox 2012 (Chapter 1); Kooken et al [manuscript in preparation, Chapter 2). Currently biochemical tests, (e.g. Staph-Ident strips) are the primary methods used for *Staphylococcus* species identification. However it is widely accepted that biochemical tests do not accurately speciate CoNS (Rhoden and Miller, 1995). As documented in the previous chapter, species identification using variation in sequence of protein markers correlates well with variation in DNA sequence of the sodA gene. The sodA gene has been most widely used in genetic speciation of CoNS but it has only been validated recently by an independent approach (Chapter 2).

Bands migrating at approximately 100 kD were separated using 1D gel electrophoresis. Profiling of peptides, after tryptic digestion, of isolated 100kDa proteins allowed staphylococcal speciation. These bands were selected because there are relatively few proteins present at this high MW simplifying separation from other proteins and larger proteins potentially contain more sequence information (Fox et al., 2011). Matrix assisted time of ionization/desorption mass spectrometry and tandem mass spectrometry (MALDI TOF MS and MALDI TOF MS and MS/MS) or LC-MS-MS analysis of gel bands identified aconitate hydratase and/or oxoglutarate dehydrogenase as the dominant proteins present in these gel bands.

Alternatively, digests of proteins released from whole cells can be converted to tryptic peptides for the determination of species (Chapter 2). In this chapter (3), the proteomic and bioinformatics technology is described in greater detail and the validity of various biomarkers evaluated. LC-MS-MS of whole cell proteins eliminated time-consuming steps involving 1D gel separation and protein extraction from the gel slices. Numerous proteins were identified based on peptide sequences and evaluated in identification of species of *Staphylococcus*. In top down proteomics, tandem mass spectrometry of whole proteins is employed to provide large scale characterization of an organism with sequenced genomes (Kelleher, 2004). In an attempt to increase accuracy in species identification and simplify sample preparation a bottom-up proteomics approach was taken here employing tryptic peptides.

The current research describes a method in of peptide analysis for the whole cell supernatant using LC-ESI MS/MS. The bottom-up proteomics approach usually utilizes instruments such as the LTQ ion trap or with greater confidence, the LTQ-Orbitrap, due

to its ability to obtain high resolution in molecular weight analysis of parent ions followed my MS-MS performed with an ion trap for fast scanning of product ions (Hu et al., 2005; Macek et al., 2006); this approach was used here for identification of *Staphylococcus* species. The current report is among the first and most definitive studies using sequence variation in newly discovered protein markers for bacterial speciation. This method was applied to a total of 29 strains including 10 CoNS species and *S. aureus*.

After LC-MS-MS analysis, bacterial speciation was performed in an automated fashion comparing experimental spectra versus virtual spectra using the software X!Tandem. The top hit or greatest confidence in protein identification (usually elongation factor tu or enolase) were found to provide accurate speciation. Bioinformatics was performed employing the widely used software X!Tandem which uses sequenced genomes to generate a virtual peptide database for comparison to experimental data. The search database was modified to include the genomes of the 11staphylococcal species most commonly isolated from man. Previous species markers of aconitate hydratase and oxoglutarate dehydrogenase (identified from gel bands) were also compared for significance versus the newly described markers identified derived from whole cell supernatants. The methodological approach described can have applications for bacterial identification across multiple species

4.2 **Materials and Methods**

4.2.1 ***Strains analyzed***

aso2c21, aso15c28, aso15c40y, aso15c106, *S. aureus* ATCC 12598, *S. chonii* ATCC 29972, *S. chromogenes* ATCC 43764, CNS20, cow924RR, cow970RR, mus5949, mus5951, 09-304-034, *S. capitis* ATCC 27841, CNS5, *S. epidermidis* ATCC12228, 900-200-150, 990rr, aso1c8, aso2c53, aso2c63, aso3c6, aso3c59w, aso15c64, aso15c84, asuno2-15c6y, asuno15c6y, CNS 1, CNS18

4.2.2 ***Culture conditions and sample preparation***

Bacteria were grown on nutrient agar plates at 37 °C for 24-48 h. The isolates were tentatively identified as staphylococci by Gram stain morphology and glucose fermentation. Colonies were harvested from plates using 0.1 M NaCl, 50 mM Tris HCl, 0.5 mM phenylmethylsulfonyl fluoride [PMSF] and placed in a FastPrep®-24 (MP Biomedicals, Solon, OH) for 6 ms X 30 sec with 5 min on ice between each cycle for a total of 6 cycles. The samples were then centrifuged at 4 °C for 1 h at 10,000 g. The supernatants were removed and placed at -70 °C for two freeze-thaw cycles to eliminate DNA.

4.2.3 ***Protein separation, tryptic digestion and MS analysis***

Bacteria were harvested (after growing as confluent lawns) from plates using 0.1 M NaCl, 50 mM Tris HCl, 0.5 mM phenylmethylsulfonyl fluoride [PMSF] and placed in a FastPrep®-24 bead beater (MP Biomedicals, Solon, OH) for 6 ms X 30 sec with 5 min on ice between each cycle for a total of 6 cycles to release proteins. The samples were then centrifuged at 4 °C for 1 h at 10,000 *g*. The supernatants were removed and placed at -70 °C for two freeze-thaw cycles to eliminate DNA.

4.2.4 ***Peptide Preparation***

Bacteria samples were thawed, vortexed briefly to re-suspend. Fifty µl of supernatant was transferred to labeled low protein binding 1.5ml microfuge tubes. 50 µl of freshly made 8M urea, 1µl of ƃ-mercapto ethanol, 24 ul of water, and 25 µl of 200 mM Tris-HCl pH 8.0 were added tubes were vortexed briefly and incubated at 60°C for one hour in a Thermomixer (ThermoFisher Scientific) shaking at 300RPM. The tubes were centrifuged briefly after the incubation to collect the evaporate on top of the lids. 800ul of 50mM ammonium bicarbonate was added to each tube to reduce the urea concentration to below 1M. 2 µl of trypsin gold at a concentration of 2 µg/µl was added to each tube and briefly vortexed to mix. The samples were then incubated at 37°C for 15 hours in a Thermomixer shaking at 300RPM.

4.2.5 ***Purification of peptides***

Solid phase extraction (SPE) was performed with a vacuum manifold using Strata C-18 T solid phase extraction columns (Phenomenex, Torrence, CA) and following the manufacturerôs protocol. Briefly, 1 ml of 100% methanol was added to activate the resin, followed by a conditioning step of 1ml 0.1% TFA water, then addition of the samples. The samples are washed with 5% acetonitrile in 0.1% TFA water, and finally elution of the samples with 80% acetonitrile in 0.1% TFA water into labeled clean low protein binding 1.5ml microfuge tubes. The desired flow rate for vacuuming steps is 0.5ml/min with the vacuum seal released after each solution. Samples were dried down to near completeness (5-10 µl remaining) with a Thermo speed vac. 25ul of 0.5% formic acid was added to each sample, using the pipettor gently wash the sides of the tube to recover as much of the sample as possible. The samples were then transferred to labeled HPLC vials with 200 µl inert glass inserts and capped with screw caps.

4.2.6 ***MS-MS analysis of the peptides***

Peptides were separated using an Agilent 1200 HPLC with a 40cm long 0.15 mm ID fused silica column packed with Jupiter 5µm C-18 resin. Column was made in house. A 50 min gradient was established by changing the relative concentrations of a two solvent systems where ñAòis 5% acetonitrile, 0.1% formic acid in H2O and ñBòis 95%

acetonitrile, 0.1% formic acid in H2O at a flow rate of 2 µl per minute. The separation had a 10 min isocratic step at 5% B and a gradient from 5% to 60% B over 50 minutes. Eluate from the HPLC was directly transferred into an LTQFT Orbitrap system (Thermo electron, Billercia MA). The electrospray conditions used were: 3.5kV spray voltage, 200°C and 200V ion transfer tube voltage. The ion injection time was set for automatic gain control with a maximum injection time of 200ms for 5X10e7 charges in the trap. The MS parent scan was performed using the Orbitrap mass analyzer using a resolution setting of 30,000. Dynamic parent ion selection was performed where the top five most abundant ions were selected for MS-MS in the linear quadrupole ion trap using a 3 m/z mass window.

4.2.7 ***Database searches***

Searches employing MS/MS data were performed using the open source software X!Tandem (www.thegpm.org/tandem)(Craig et al., 2004). Raw files from the Orbitrap (MS-MS data) were converted into the mascot general format (msg) files) using the program Proteo-wizard (*proteowizard.sourceforge.net)*. The data base was modified by downloading the genomes from all eleven species representing the 29 strains in this study. Fasta files were downloaded from (Uniprot, www.uniprot.org). and then converted to FastaPro files for use with X!Tandem. The Fasta-pro files were placed in the fasta section of the software the pro_species.jl (controls program-interface for bacterial species) and taxonomy.xml files (lists each file pathway for each fastaPro file)

of the X!Tandem software. Trypsin-specific enzymatic digestion rules were selected and two missed cleavages. The parent mass accuracy was specified at +/- 10 ppm (data acquired with Orbitrap mass spectrometer) and fragment mass accuracy of +/- 300 ppm for MS-MS scans acquired with the LTQ system. Parent mass error was set to 10 parts per million (PPM), maximum parent charge set to 5, and no modifications selected for these samples. The X!Tandem output was also processed into a file format compatible with the mass spectrometry generating function (MSGF) for secondary validation (Kim et al., 2008). Peptide spectrum matches with e-values less than -100 were determined using the MSGF were excluded from consideration unless noted otherwise. All peptides matches were captured into a SQLite database for sorting of matches.

Species and strains in custom database: Staphylococcus aureus O46, *Staphylococcus aureus* strain Newman, *Staphylococcus aureus* subsp *aureus* CIG1242, *Staphylococcus capitis*, *Staphylococcus capitis* SK14, *Staphylococcus capitis* subsp *capitis*, *Staphylococcus capitis* subsp *urealyticus*, *Staphylococcus capitis* VCU116, *Staphylococcuscarnosus*, *Staphylococcus carnosus* strain TM300, *Staphylococcus carnosus* subsp *carnosus*, *Staphylococcus carnosus* subsp *utilis*, *Staphylococcus chromogenes*, *Staphylococcus cohnii*, *Staphylococcus cohnii* subsp *cohnii*, *Staphylococcus cohnii* subsp *urealyticus*, *Staphylococcus epidermidis* strain ATCC 12228, *Staphylococcus epidermidis* strain ATCC 35984 RP62A, *Staphylococcus epidermidis* VCU127, *Staphylococcus haemolyticus*, *Staphylococcus haemolyticus* strain JCSC1435, *Staphylococcus hominis*, *Staphylococcus hominis* SK119, *Staphylococcus hominis* subsp *hominis*, *Staphylococcus hominis* subsp *hominis* C80,

Staphylococcus hominis subsp *novobiosepticus*, *Staphylococcus hominis* VCU122, *Staphylococcus lugdunensis*, *Staphylococcus lugdunensis* ACS-027 V Sch2, *Staphylococcus lugdunensis* M23590, *Staphylococcus lugdunensis* strain HKU09-01, *Staphylococcus lugdunensis* strain N920143, *Staphylococcus lugdunensis* VCU139, *Staphylococcus saprophyticus*, *Staphylococcus saprophyticus* subsp *bovis*, *Staphylococcus saprophyticus* subsp *saprophyticus*, *Staphylococcus saprophyticus* subsp *saprophyticus* KACC 16562, *Staphylococcus saprophyticus* subsp *saprophyticus* MS1146, *Staphylococcus saprophyticus* subsp saprophyticus strain ATCC 15305 DSM 20229, *Staphylococcus simulans* ACS-120-V-Sch1, *Staphylococcus warneri*, *Staphylococcus warneri* L37603, *Staphylococcus warneri* VCU121

4.3 **Results and Discussion**

Our previous method of MS and MS/MS bacterial identification employed SDS-PAGE gels for protein separation along with the additional process of excising the band of interest and extracting the protein from the gel slice (Fox et al., 2011, Kooken, Fox and Fox 2012 and Chapter 2). Overall analysis from start to finish takes approximately 5 days. By utilizing whole cell trypsin digests the sample preparation time was decreased to one day. Furthermore when using whole cell digests multiple proteins may be identified leading to greater discrimination in species identification.

MS scans were performed on the Orbitrap, where the 50 most abundant peaks per scan were then sent on for MS-MS analysis in the ion trap. This setup allows for

high resolution and fast initial MS scans, followed by a MS-MS scan that have a more focused MW range and fast scan times. However resolution is limited by the use of ion trap. Since there are many peptides that correspond to multiple proteins analyzed with a whole cell preparation the ability to identify proteins that have been extensively studied and annotated within a database increases.

MS/MS results were analyzed using X!Tandem search engine in which genomic sequence information gathered from the UNIPROT data base for the 11 most common *Staphylococcus* species associated with humans was uploaded. The X!Tandem program allows for additions of genomic sequences to what is already provided in the standard database so a more complete evaluation of spectra is possible. This customized library was instigated using genomic sequences stored in the UNIPROT database. These species and strains were aided by previous work from our group studying staphylococci released from human skin and collected from indoor air in crowded school rooms (Fox et al 2005, Fox et al., 2008, Fox et al., 2010; Fox et al., 2011). To prevent over sampling of one species, such as *S. aureus* the genomic sequences for the highest reported, and fully annotated strains were downloaded into X!Tandem. With all *Staphylococcus* species selected the advance settings on the program allowed for several adjustments to be made. The results from X!Tandem give an individual score for probability of each protein identification, this is noted a as the E-Score in both Tables 4.1 and 4.2. The smaller the score the more peptides matched up to the protein sequence. Values above 100 are considered highly significant.

In utilizing whole cell supernatants only the most abundant peptides are readily detected, showing up in multiple scans and corresponding to peptides expected in the

customized X!Tandem database. In all instances the most frequently identified peptide for whole supernatant with the best coverage was elongation factor tu, and enolase (see Table 4.1). These proteins are also useful because they have been studied extensively in *Staphylococcus* species and information on the amino acid sequences has been annotated for even lesser studied species. Selected ATCC type strains were run as controls to verify proper identification, top results *for S. aureus, S. chonii, S. chromogenes, S. capitis*, and *S. epidermidis* being elongation factor tu, and its e-score markedly higher than the next given species further supported the hypothesis that not only can whole supernatant be used to identify important proteins, but there is good enough coverage of both unique and conserved regions for proper species identification. These controls also assisted in evaluating the use of aconitate hydratase and oxoglutarate dehydrogenase for speciation in a whole protein supernatant (Table 4.2).

It was found that both aconitate hydratase and oxoglutarate dehydrogenase could only be identified in 2 of the 5 ATCC strains analyzed (*S. epidermidis*, and *S. capitis*). This pattern of inability to successfully identify these proteins with coverage of both unique and conserved regions was also displayed with the unknown samples. Fig. 4.1 displays the results for stain aso2c21, as a demonstration of peptide coverage the majority of the sequence for elongation factor tu in *S. aureus* has been positively identified from daughter ions in the sample and is displayed as red letters indicating the amino acid sequence has been successfully matched. In this figure the coverage allows positive identification for not only the protein, but the species to which it belongs. Coverage with peptides detected derived from whole cell supernatant covers the

majority of the protein's sequence. In contrast on evaluating aconitate hydratase and oxoglutarate dehydrogenase as markers the results indicated scarcely any coverage, and often times the peptides that did match were conserved across multiple species (Fig. 4.2).

It was found that aconitate hydratase and oxoglutarate dehydrogenase, when isolated and analyzed as an individual protein in previous work resulted in exceptional coverage of all regions and allowed for species identification, but in whole cell supernatant these proteins are not abundant enough for detection. In the whole cell supernatant aconitate hydratase was detected in 19 of the 29 samples, and with a minimum acceptable e-value of -20 (value at which at least 4 peptides are positively matched to the genomic sequence) only 15 of these scored high enough for a significant match (Table 4.2). Oxoglutarate dehydrogenase showed up 16 times but only in *S. capitis* ATCC 27841, *S. epidermidis* ATCC 12228, and 09-304-034 were the scores above -20 and had multiple peptides match.

In comparison for all of the samples elongation factor tu, enolase, or both were found to be top hits, with close to full coverage even on unique regions which are valuable in differentiating amongst the species (Table 4.1). Since we are not limiting ourselves to one particular protein the top score is whatever protein has the most number of correct matches; the better the coverage the more accurate a match. This technique has been able to identify species with greater accuracy then gel slices due to the broader view of the organism. The matches from whole cell protein analysis agreed with sodA gene sequence variation (see Chapter 2).

4.0 **Conclusions**

Looking at all proteins expressed, instead of a pre-determined one (e.g. found in a gel slice)allows more options for positive identification of an organism. Previous work focused on a top down approach, starting with 1D SDS PAGE gel slices of the proteins aconitate hydratase and oxoglutarate dehydrogenase for use as marker proteins. The main drawback with utilizing these proteins successfully in whole cell supernatant samples is that if the protein(s) selected is not abundant then accurate identification with multiple peptides is close to impossible. Therefore, by looking at whichever protein provides the best match, the largest numbers of matching peptides are also selected; this allows for consistent and accurate speciation.

This methodology, although successful, is still in its infancy and with further refinement it may have implications for clinical speciation in the future at the current time development of a custom-data base, along with expensive equipment, such at the Orbitrap, requires expert knowledge which may prove to be prohibitive in clinical microbiology settings. With future development and implementation of user-friendly software and instrumentation the knowledge gained from this research may be applied for successful biologic sample identification. MALDI profiling itself took almost 20 years to develop in clinical microbiology settings, and with future advances and simplification whole cell protein analysis has the same potential (Intellicato and Fox, 2013). In research settings, LC-MS-MS may help expand the knowledge-base of abundant

proteins (with unique regions) derived from bacterial species allowing their identification or detection in simple or complex matrices.

Table 4.1 Protiens and species identified with the highest confidence

	Rank	Protein	Score	Species
aso2c21	1	Elongation factor tu	-459.50	*Staphylococcus warneri L37603*
	2	Elongation factor G	-209.50	*Staphylococcus warneri VCU121*
	3	Enolase	-175.30	*Staphylococcus warneri L37603*
aso15c28	1	Elongation factor tu	-338.20	*Staphylococcus haemolyticus (strain JCSC1435)*
	2	Elongation factor tu	-239.90	*Staphylococcus epidermidis (strain ATCC 12228)*
	3	Formate--tetrahydrofolate ligase	-126.00	*Staphylococcus haemolyticus (strain JCSC1435)*
aso15c40y	1	Elongation factor tu	-102.5	*Staphylococcus carnosus (strain TM300)*
	2	Translation elongation factor tu	-59.5	*Staphylococcus haemolyticus*
	3	Enolase	-49.3	*Staphylococcus epidermidis (strain ATCC 12228)*
aso15c106	1	Elongation factor tu	-366	*Staphylococcus haemolyticus (strain JCSC1435)*
	2	Elongation factor tu	-315.2	*Staphylococcus warneri L37603*
	3	Formate--tetrahydrofolate ligase	-123.2	*Staphylococcus haemolyticus (strain JCSC1435)*
S. aureus ATCC12598	1	Elongation factor tu	-276.2	*Staphylococcus aureus O46*
	2	Enolase	-237.2	*Staphylococcus aureus O46*
	3	Elongation factor G	-161	*Staphylococcus aureus O46*
S. chonii ATCC29972	1	Elongation factor tu	-255.7	*Staphylococcus cohnii subsp. cohnii*
	2	Elongation factor tu	-199.5	*Staphylococcus saprophyticus (KACC 16562)*
	3	Elongation factor tu	-160.6	*Staphylococcus hominis SK119*
S. chromogenes ATCC43764	1	Elongation factor tu	-208.8	*Staphylococcus chromogenes*
	2	Elongation factor tu	-156.8	*Staphylococcus capitis SK14*
	3	Translation elongation factor tu	-116.3	*Staphylococcus haemolyticus*
CNS20	1	Elongation factor tu	-299.4	*Staphylococcus aureus O46*
	2	Enolase	-251.6	*Staphylococcus aureus O46*
	3	Elongation factor G	-209.6	*Staphylococcus aureus O46*
cow924RR	1	Elongation factor tu	-169.2	*Staphylococcus capitis SK14*
	2	Elongation factor tu	-118.4	*Staphylococcus simulans ACS-120-V-Sch1*
	3	60 kDa chaperonin (GroEL)	-88.7	*Staphylococcus saprophyticus (KACC 16562)*
cow970RR	1	Elongation factor tu	-400.8	*Staphylococcus saprophyticus (KACC 16562)*

	2	Enolase	-159.7	*Staphylococcus saprophyticus (KACC 16562)*
	3	1-pyrroline-5-carboxylate dehydrogenase	-106.9	*Staphylococcus saprophyticus (KACC 16562)*
mus5949	1	Elongation factor tu	-296.1	*Staphylococcus chromogenes*
	2	Elongation factor tu	-113.9	*Staphylococcus capitis SK14*
	3	Glyceraldehyde-3-phosphate dehydrogenase	-110.6	*Staphylococcus chromogenes*
mus5951	1	Elongation factor tu	-150.4	*Staphylococcus warneri L37603*
	2	Nucleoid DNA-binding protein HU	-113.5	*Staphylococcus saprophyticus (KACC 16562)*
	3	60 kDa chaperonin (GroEL)	-102.3	*Staphylococcus saprophyticus (KACC 16562)*
09-304-034	1	Elongation factor tu	-315.7	*Staphylococcus saprophyticus (KACC 16562)*
	2	Enolase	-212.1	*Staphylococcus saprophyticus (KACC 16562)*
	3	Elongation factor tu	-179.6	*Staphylococcus hominis SK119*
S. capitis ATCC27841	1	Elongation factor tu	-357.9	*Staphylococcus capitis SK14*
	2	Elongation factor tu	-261.9	*Staphylococcus epidermidis (strain ATCC 12228)*
	3	Elongation factor G	-218	*Staphylococcus capitis SK14*
CNS5	1	Elongation factor tu	-371.4	*Staphylococcus simulans ACS-120-V-Sch1*
	2	Elongation factor tu	-313.8	*Staphylococcus lugdunensis ACS-027-V-Sch2*
	3	Elongation factor tu	-238	*Staphylococcus capitis SK14*
S. epidermidis ATCC12228	1	Elongation factor tu	-441.7	*Staphylococcus epidermidis (strain ATCC 12228)*
	2	Elongation factor tu	-254.2	*Staphylococcus hominis SK119*
	3	Elongation factor G	-231.1	*Staphylococcus epidermidis (strain ATCC 12228)*
900-200-150	1	Elongation factor tu	-344.9	*Staphylococcus warneri L37603*
	2	Isocitrate dehydrogenase [NADP]	-136.8	*Staphylococcus warneri L37603*
	3	Elongation factor G	-127.2	*Staphylococcus warneri VCU121*
990rr	1	Elongation factor tu	-245.6	*Staphylococcus simulans ACS-120-V-Sch1*
	2	Elongation factor G	-173.7	*Staphylococcus simulans ACS-120-V-Sch1*
	3	Citrate synthase	-137.2	*Staphylococcus simulans ACS-120-V-Sch1*
aso1c8	1	Elongation factor tu	-243.1	*Staphylococcus aureus O46*
	2	Enolase	-206.1	*Staphylococcus aureus O46*
	3	Enolase	-137	*Staphylococcus hominis SK119*
aso2c53	1	Elongation factor tu	-353.5	*Staphylococcus epidermidis (strain ATCC 12228)*
	2	Elongation factor G	-159.8	*Staphylococcus epidermidis (strain ATCC 12228)*
	3	Elongation factor tu	-138.8	*Staphylococcus haemolyticus*
aso2c63	1	Elongation factor tu	-347.2	*Staphylococcus epidermidis (strain ATCC 12228)*
	2	Elongation factor tu	-170.9	*Staphylococcus haemolyticus*
	3	Enolase	-152.1	*Staphylococcus epidermidis (strain ATCC 12228)*
aso3c6	1	Enolase	-230.2	*Staphylococcus aureus O46*
	2	Elongation factor tu	-215.1	*Staphylococcus aureus O46*
	3	Elongation factor tu	-147.2	*Staphylococcus haemolyticus (strain JCSC1435)*
aso3c59w	1	Elongation factor tu	-277.9	*Staphylococcus lugdunensis ACS-027-V-Sch2*
	2	Elongation factor tu	-251.3	*Staphylococcus warneri L37603*
	3	DNA-binding protein II	-129.9	*Staphylococcus aureus O46*
aso15c64	1	Enolase	-226.8	*Staphylococcus aureus O46*
	2	Elongation factor tu	-205.1	*Staphylococcus aureus O46*
	3	Elongation factor G	-173.9	*Staphylococcus aureus O46*
aso15c84	1	Enolase	-217.6	*Staphylococcus aureus O46*
	2	Elongation factor tu	-213	*Staphylococcus aureus O46*
	3	Alkaline shock protein 23	-140.3	*Staphylococcus aureus O46*

asuno2-15c6y	1	Elongation factor tu	-185.9	*Staphylococcus aureus O46*
	2	Enolase	-167.6	*Staphylococcus aureus O46*
	3	Elongation factor G	-123	*Staphylococcus aureus O46*
asuno15c6y	1	Enolase	-270.2	*Staphylococcus aureus O46*
	2	Elongation factor tu	-227.9	*Staphylococcus aureus O46*
	3	Elongation factor G	-125.9	*Staphylococcus aureus O46*
CNS 1	1	Elongation factor tu	-219.6	*Staphylococcus simulans ACS-120-V-Sch1*
	2	Uncharacterized protein	-88.2	*Staphylococcus simulans ACS-120-V-Sch1*
	3	Uncharacterized protein	-69.8	*Staphylococcus simulans ACS-120-V-Sch1*
CNS18	1	Elongation factor tu	-227.3	*Staphylococcus aureus O46*
	2	Enolase	-145.4	*Staphylococcus aureus O46*
	3	Ornithine--oxo-acid transaminase	-123.1	*Staphylococcus aureus O46*

log(e): the base-10 log of the expectation that any particular protein assignment was made at random (E-value)

Table 4.2 Confidence levels aconitate hydratase and oxoglutarate dehydrogenase

Sample	Rank	Protein	Score	Species
900-200-150	44	Aconitate hydratase	-28	*Staphylococcus warneri L37603*
	37	oxoglutarate dehydrogenase	-32.1	*Staphylococcus warneri L37603*
09-304-034	7	Aconitate hydratase	-93.8	*Staphylococcus saprophyticus (KACC 16562)*
mus5951	16	Aconitate hydratase	-44.9	*Staphylococcus saprophyticus (KACC 16562)*
	133	oxoglutarate dehydrogenase	-1.8	*Staphylococcus saprophyticus (KACC 16562)*
990rr	34	oxoglutarate dehydrogenase	-19.6	*Staphylococcus simulans ACS-120-V-Sch1*
cow924RR	12	Aconitate hydratase	-50.8	*Staphylococcus saprophyticus (KACC 16562)*
cow970RR	9	Aconitate hydratase	-85.5	*Staphylococcus saprophyticus (KACC 16562)*
	66	oxoglutarate dehydrogenase	-9.6	*Staphylococcus saprophyticus (KACC 16562)*
CNS5	14	Aconitate hydratase	-75.2	*Staphylococcus simulans ACS-120-V-Sch1*
	175	oxoglutarate dehydrogenase	-1.3	*Staphylococcus simulans ACS-120-V-Sch1*
CNS18	20	Aconitate hydratase	-37.1	*Staphylococcus aureus O46*
S. capitis ATCC27841	22	Aconitate hydratase	-56.3	*Staphylococcus capitis SK14*
	34	oxoglutarate dehydrogenase	-44.6	*Staphylococcus capitis SK14*
S. epidermidis ATCC12228	63	Aconitate hydratase	-25.6	*Staphylococcus epidermidis (strain ATCC 12228)*
	70	oxoglutarate dehydrogenase	-23.5	*Staphylococcus epidermidis (strain ATCC 12228)*
aso2c21	22	Aconitate hydratase	-56.00	*Staphylococcus warneri L37603*
	96	oxoglutarate dehydrogenase	-14.40	*Staphylococcus warneri L37603*
aso1c8	48	Aconitate hydratase	-21.8	*Staphylococcus aureus O46*
	53	oxoglutarate dehydrogenase	-19.6	*Staphylococcus aureus O46*
aso2c21	22	Aconitate hydratase	-56.00	*Staphylococcus warneri L37603*
aso2c53	79	Aconitate hydratase	-5.5	*Staphylococcus epidermidis (strain*

				ATCC 12228)
aso2c63	34	Aconitate hydratase	-21.7	*Staphylococcus epidermidis (strain ATCC 12228)*
	98	oxoglutarate dehydrogenase	-2.7	*Staphylococcus epidermidis (strain ATCC 12228)*
aso3c6	37	Aconitate hydratase	-27.5	*Staphylococcus aureus O46*
	104	oxoglutarate dehydrogenase	-2.5	*Staphylococcus aureus O46*
aso15c64	46	Aconitate hydratase	-17.2	*Staphylococcus aureus O46*
	99	oxoglutarate dehydrogenase	-3.6	*Staphylococcus aureus O46*
aso15c84	51	Aconitate hydratase	-20.9	*Staphylococcus aureus O46*
	112	oxoglutarate dehydrogenase	-4.8	*Staphylococcus aureus O46*
asuno2-15c6y	48	Aconitate hydratase	-16.9	*Staphylococcus aureus O46*
	117	oxoglutarate dehydrogenase	-2	*Staphylococcus aureus O46*
asuno15c6y	20	Aconitate hydratase	-40.7	*Staphylococcus aureus O46*
	95	oxoglutarate dehydrogenase	-4.5	*Staphylococcus aureus O46*

log(e): the base-10 log of the expectation that any particular protein assignment was made at random (E-value)

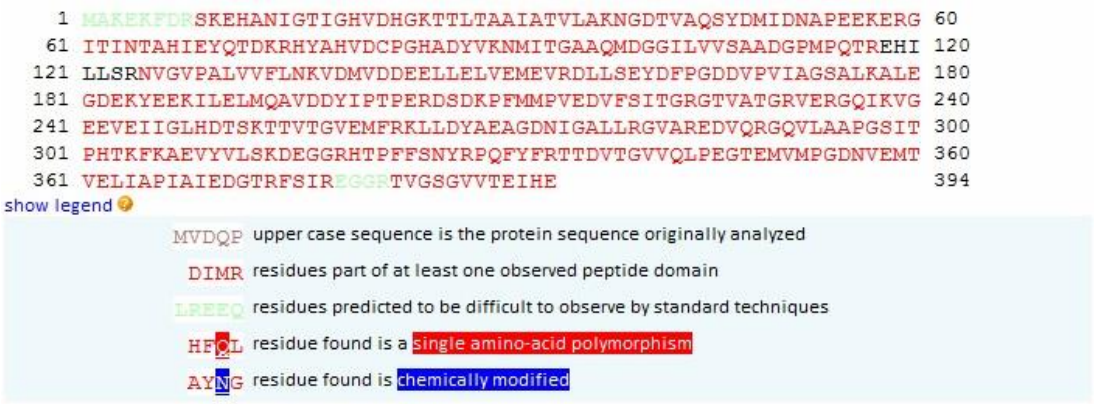


Figure 4.1 aso2c21: elongation factor tu sequence, and matching peptides from sample in red letters.

Figure 4.2 Aso2c21 amino acids in red letters match up with sequence for oxoglutarate dehydrogenase protein

CHAPTER 5

CONCLUSIONS

5.0 Taxonomic markers to identify *Staphylococcus* genus from co-contaminants

Beginning with the characterization of natural flora found on human skin it was found that there were abundant levels of *Micrococcus* and *Staphylococcus* present in collections from environmental settings (chapter 2). Work performed by others previous to this dissertation found classification of these two generas difficult with available tests methods (chapter 2). It was even shown that strains the FDA used in testing were not properly identified, and in fact one of the standards used was not the *Micrococcus* species it was labeled as but instead a closely related genus(Tang, 2003). When researched was found previously that over two thirds of the environmental isolates were believed to be *Micrococci* (Bruker Biotyping), it created a need for greater accuracy in characterization (Satta et al., 1993). Tryptic peptide analysis of the aconitate hydratase protein with MALDI TOF MS established criteria to confirm species identity for the environmental isolates. The more in depth technique of MALDI TOF-TOF MS-MS then was used to confirm protein identity, and also show a clear difference between the profiles of the two deposited *M. luteus*, and how it differs from *Staphylococcus* (chapter 2). This work created the framework to allow *Micrococcus*, a harmless co-contaminant in environmental and biologically derived samples to be differentiated from

Staphylococcus, which can now be further studied for information on species identification.

After establishing a means to accurately identify the *Staphylococcus* genus speciation amongst *Staphylococcal* species proved to be a complex issue, with many species observed suspected to be not identified or inaccurately specieated.

5.1 **Development of 'bottom-up' approach for *Staphylococcus* speciation**

The 3rd chapter focused on a means to properly speciate *Staphylococcus* isolates by analyzing protein sequence variations. Tandem mass spectrometry performed on trypsin-released peptides gave amino acid sequences of both conserves and unique regions within the aconitate hydratase and oxoglutarate dehydrogenase proteins. By looking at individual proteins and peptide matches it was possible to establish unique sequences within the protein that can be used to identify one species from another. The MS-MS results that provided the information of protein and species identity correlated well with sodA sequencing from microarrays. The microarrays allowed all staphylococcal species to be differentiated simultaneously with their own sequence specific probe (based on previously discovered sequence variation) and were used as a means to further verify the MS-MS method could correctly identify species without knowledge of sample identity. From this it was shown that tandem mass spectrometry is able to discover and use new protein marker sequences that can be identified as the need arises.

5.2 **LC-ESI MS-MS use for whole cell protein analysis allows the fastest and most accurate means for CoNS species identification**

The developmental work described in the 4th chapter is a significant step forward in the process of reliable speciation of CoNS specimens. Tandem mass spectrometry, demonstrated the ability to look at new protein marker sequences to be used for species identification as the need arises. By looking at all proteins expressed, instead of a pre-determined ones with gel separation a better picture of the identity of an organism is provided, along with more options for positive species identification. This whole cell method is a refinement from previous work which focused on a top down approach by protein selection of hydratase and oxoglutarate dehydrogenase as marker proteins. A reduction in time due to the simplified materials processing procedure was also established. It was found that the use of these proteins is only successful as whole cell supernatant samples if the protein(s) selected is/are abundant. Previously studied aconitate hydrates, and oxoglutarate dehydrogenase were often not abundant enough study without purification (table 4.2) so in order to remedy this in whole supernatant evaluation occurs of whichever protein provides the best match and the largest numbers of matching peptides. This delivered consistent and accurate speciation across all samples.

5.3 **Evaluation and future directions for proteomic species identification**

Proteomics is largely applied to cataloging protein expression with little emphasis on taxonomic identification. This dissertation focused on identifying CoNS species in order to introduce methodology in which mass spectrometry may be used for bacterial species identification. The need for improved taxonomic identification stems from current techniques failing in speciation of closely related isolates. This often arises when some species are thought to be clinically significant and other are ignored, such as most CoNS species were believed to be incapable of human disease. The discovery of multiple CoNS species causing infections in clinical settings, coupled with the fact that even previously found clinically significant *Staphylococcus* species are often misidentified created a need for consistent and accurate taxonomic classification. With the failure of physiological and molecular tests to speciate CoNS, it was found that better but sometimes inaccurate species identification was possible with basic MALDI-TOF MS (Fox et al., 2011). Even though MALDI TOF MS for bacterial speciation, was described by several groups close to 20 years ago utilization of this technique has only come into clinical microbiology labs in the past few years (Intelicato-Young and Fox, 2013). Even with this advance from physiological tests MALDI TOF MS is not always capable of the desired specificity for the complex issue of staphylococcal speciation. In which case a more in depth approach, such as MS-MS proves to provide the consistent and accurate results for species specific identification/information. At this time utilization of tandem mass spectrometry can provide the greatest amount of information with little known about a sample and has promise for success across multiple species.

This methodology is still in its infancy and although it may have implications for clinical speciation in the future at the current time generation of a custom-data base along with expensive equipment that requires expert knowledge to run (e.g. the Orbitrap) make it prohibitive for use in clinical microbiology settings. However, it is optimistic that this situation could change in the future with implementation of user-friendly software and instrumentation. Indeed it has taken almost 20 years for the much simpler MALDI profiling to be fully developed and accepted by the clinical microbiology community (Intellicato and Fox, 2013). In research settings, LC-MS-MS may help expand the knowledge-base of abundant proteins (with unique regions) derived from bacterial species allowing their identification or detection in simple or complex matrices. Further studies could evaluate its use in species identification across samples from different origins and genus, testing the established methods to differentiate samples across many taxonomic levels.

WORKS CITED:

Aebersold, R. and Mann, M., 2003. Mass spectrometry-based proteomics. Nature, 422: 198-207.

Angel, T.E. et al., 2012. Mass spectrometry-based proteomics: existing capabilities and future directions. Chem Soc Rev, 41(10): 3912-28.

Baird-Parker, A.C., 1965. Staphylococci and their classification. Ann N Y Acad Sci, 128: 4-25.

Baker, J.S., 1984. Comparison of Various Methods for Differentiation of Staphylococci and Micrococci. Journal of Clinical Microbiology, 19: 875-879.

Baker, M., 2010. Mass spectrometry for biologists. Nat Meth, 7: 157-161.

Baldellon, C. and Megraud, F., 1985. Characterization of Micrococcaceae strains isolated from the human urogenital tract by the conventional scheme and a micromethod. J Clin Microbiol, 21: 474-7.

Billeci, T.M. and Stults, J.T., 1993. Tryptic mapping of recombinant proteins by matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem, 65: 1709-16.

Blaum, K., 2006. High-accuracy mass spectrometry with stored ions. Physics Reports, 425: 1-78.

Brodie, E.L. et al., 2007. Urban aerosols harbor diverse and dynamic bacterial populations. Proc Natl Acad Sci U S A, 104: 299-304.

Canas, B., Lopez-Ferrer, D., Ramos-Fernandez, A., Camafeita, E. and Calvo, E., 2006. Mass spectrometry technologies for proteomics. Brief Funct Genomic Proteomic, 4: 295-320.

Carbonnelle, E. et al., 2012. Robustness of two MALDI-TOF mass spectrometry systems for bacterial identification. J Microbiol Methods, 89: 133-6.

Carbonnelle, E. et al., 2011. MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. Clin Biochem, 44: 104-9.

Cash, P., 2009. Proteomics in the study of the molecular taxonomy and epidemiology of bacterial pathogens. ELECTROPHORESIS, 30: S133-S141.

Center, K.J., Reboli, A.C., Hubler, R., Rodgers, G.L. and Long, S.S., 2003. Decreased vancomycin susceptibility of coagulase-negative staphylococci in a neonatal intensive care unit: evidence of spread of Staphylococcus warneri. J Clin Microbiol, 41: 4660-5.

Chowdhury, S.K., Katta, V. and Chait, B.T., 1990a. An electrospray-ionization mass spectrometer with new features. Rapid Commun Mass Spectrom, 4: 81-7.

Chowdhury, S.K., Katta, V. and Chait, B.T., 1990b. Electrospray ionization mass spectrometric peptide mapping: a rapid, sensitive technique for protein structure analysis. Biochem Biophys Res Commun, 167: 686-92.

Cornish, T.J. and Cotter, R.J., 1993. A curved-field reflectron for improved energy focusing of product ions in time-of-flight mass spectrometry. Rapid Commun Mass Spectrom, 7: 1037-40.

Craig, R., Cortens, J.P. and Beavis, R.C., 2004. Open source system for analyzing, validating, and storing protein identification data. J Proteome Res, 3: 1234-42.

Cravatt, B.F., Simon, G.M. and Yates, J.R., 3rd, 2007. The biological impact of mass-spectrometry-based proteomics. Nature, 450: 991-1000.

Davis, M.T., Stahl, D.C., Hefta, S.A. and Lee, T.D., 1995. A microscale electrospray interface for on-line, capillary liquid chromatography/tandem mass spectrometry of complex peptide mixtures. Anal Chem, 67: 4549-56.

De Paulis, A.N., Predari, S.C., Chazarreta, C.D. and Santoianni, J.E., 2003. Five-Test Simple Scheme for Species-Level Identification of Clinically Significant Coagulase-Negative Staphylococci. J Clin Microbiol, 41: 1219-1224.

Dubois, D. et al., 2010. Identification of a variety of Staphylococcus species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol, 48: 941-5.

Edmond deHoffmann, V.S., 2003. Mass Spectrometry: Principles and Applications Wiley, John & Sons, Incorporated, 502.

Evans, J.B. and Kloos, W.E., 1972. Use of shake cultures in a semisolid thioglycolate medium for differentiating staphylococci from micrococci. Appl Microbiol, 23: 326-31.

Falk, D. and Guering, S.J., 1983. Differentiation of Staphylococcus and Micrococcus spp. with the Taxo A bacitracin disk. J Clin Microbiol, 18: 719-21.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. Science, 246: 64-71.

Fox, K., Castanha, E., Fox, A., Feigley, C. and Salzberg, D., 2008. Human K10 epithelial keratin is the most abundant protein in airborne dust of both occupied and unoccupied school rooms. J Environ Monit, 10: 55-9.

Fox, K., Fox, A., Elssner, T., Feigley, C. and Salzberg, D., 2010. MALDI-TOF mass spectrometry speciation of staphylococci and their discrimination from micrococci isolated from indoor air of schoolrooms. J Environ Monit, 12: 917-23.

Fox, K., Fox, A., Rose, J. and Walla, M., 2011. Speciation of coagulase negative staphylococci, isolated from indoor air, using SDS PAGE gel bands of expressed proteins followed by MALDI TOF MS and MALDI TOF-TOF MS-MS analysis of tryptic peptides. J Microbiol Methods, 84: 243-50.

Frank, K.L., Hanssen, A.D. and Patel, R., 2004. icaA is not a useful diagnostic marker for prosthetic joint infection. J Clin Microbiol, 42: 4846-9.

Fujita, S., Senda, Y., Iwagami, T. and Hashimoto, T., 2005. Rapid identification of staphylococcal strains from positive-testing blood culture bottles by internal transcribed spacer PCR followed by microchip gel electrophoresis. J Clin Microbiol, 43: 1149-57.

Giammarinaro, P., Leroy, S., Chacornac, J.P., Delmas, J. and Talon, R., 2005. Development of a new oligonucleotide array to identify staphylococcal strains at species level. J Clin Microbiol, 43: 3673-80.

Gilad, J. and Schwartz, D., 2007. Identification of Staphylococcus species with the VITEK 2 system: the case of Staphylococcus hominis. J Clin Microbiol, 45: 685; author reply 685-6.

Goebel, B.M. and Stackebrandt, E., 1994. Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. Appl Environ Microbiol, 60: 1614-21.

Goebel, E.S.a.B.M., 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16s rRNA Sequence Analysis in the Present Species Definition in Bacteriology. Int. J. Syst. Evol. Microbiol., 44: 846-849.

Goh, S.H. et al., 1997. Identification of Staphylococcus species and subspecies by the chaperonin 60 gene identification method and reverse checkerboard hybridization. J Clin Microbiol, 35: 3116-21.

Goyal R, S.N.P., Kumar A, Kaur I, Singh M, Sunita N, Mathur M 2006. Simple and economical method for speciation and resistotyping of clinically significant coagulase negative staphylococci. Indian J Med Microbiol., 24.

Gupta, R.S., 1998. Protein Phylogenies and Signature Sequences: A Reappraisal of Evolutionary Relationships among Archaebacteria,Eubacteria, and Eukaryotes. Microbiology and Molceular Biology Reviews, 62: 1435-1491.

Hebert, G.A., Crowder, C.G., Hancock, G.A., Jarvis, W.R. and Thornsberry, C., 1988. Characteristics of coagulase-negative staphylococci that help differentiate these species and other members of the family Micrococcaceae. J Clin Microbiol, 26: 1939-49.

Henzel, W.J. et al., 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci U S A, 90: 5011-5.

Herchline, T.E. and Ayers, L.W., 1991. Occurrence of Staphylococcus lugdunensis in consecutive clinical cultures and relationship of isolation to infection. J Clin Microbiol, 29: 419-21.

Hillenkamp, F., Karas, M., Beavis, R.C. and Chait, B.T., 1991. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem, 63: 1193A-1203A.

Hirotaki, S., Sasaki, T., Kuwahara-Arai, K. and Hiramatsu, K., 2011. Rapid and accurate identification of human-associated staphylococci by use of multiplex PCR. J Clin Microbiol, 49: 3627-31.

Hu, Q. et al., 2005. The Orbitrap: a new mass spectrometer. J Mass Spectrom, 40: 430-43.

Intelicato-Young, J. and Fox, A., 2013. Mass spectrometry and tandem mass spectrometry characterization of protein patterns, protein markers and whole proteomes for pathogenic bacteria. J Microbiol Methods, 92: 381-6.

Iwase, T., Seki, K., Shinji, H., Mizunoe, Y. and Masuda, S., 2007. Development of a real-time PCR assay for the detection and identification of Staphylococcus capitis, Staphylococcus haemolyticus and Staphylococcus warneri. J Med Microbiol, 56: 1346-9.

Jabbour, R.E. et al., 2010. Double-blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics. Appl Environ Microbiol, 76: 3637-44.

Jensen, O.N., Podtelejnikov, A. and Mann, M., 1996. Delayed extraction improves specificity in database searches by matrix-assisted laser desorption/ionization peptide maps. Rapid Commun Mass Spectrom, 10: 1371-8.

Johnson, J.V., Yost, R.A., Kelley, P.E. and Bradford, D.C., 1990. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. Anal Chem, 62: 2162-2172.

Jukes, L. et al., 2010. Rapid differentiation of Staphylococcus aureus, Staphylococcus epidermidis and other coagulase-negative staphylococci and meticillin susceptibility testing directly from growth-positive blood cultures by multiplex real-time PCR. J Med Microbiol, 59: 1456-61.

Karas, M. and Hillenkamp, F., 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem, 60: 2299-301.

Kelleher, N.L., 2004. Top-down proteomics. Anal Chem, 76: 197A-203A.

Kim, S., Gupta, N. and Pevzner, P.A., 2008. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. J Proteome Res, 7: 3354-63.

Kloos, W.E., Thomas G. Tornabene, Karl H. Schliefer, 1974. Isolation and Characterization of Micrococci From Human Skin, Including Two New Species: Micrococcus lylae and Micrococcus kristinae. IJSEM 24: 79-101.

Knochenmuss, R., 2006. Ion formation mechanisms in UV-MALDI. Analyst, 131: 966-86.

Kocur, M. and Martinec, T., 1967. Contribution to the characteristics of Staphylococcus aureus. Vnitr Lek, 13: 866-71.

Kooken, J., Karen Fox, Alvin Fox, Diego Altomare, Kim Creek, David Wunschel, 2013. Identification of Staphylococcal species based on variations in protien sequence (Tandem mass spectrometry) and DNA sequence (Microarray). Manuscript in Progress.

Kwok, A.Y. et al., 1999. Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus Staphylococcus. Int J Syst Bacteriol, 49: 1181-92.

Laiko, V.V., Baldwin, M.A. and Burlingame, A.L., 2000a. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem, 72: 652-7.

Laiko, V.V., Moyer, S.C. and Cotter, R.J., 2000b. Atmospheric pressure MALDI/ion trap mass spectrometry. Anal Chem, 72: 5239-43.

Leary, J.J. and Schmidt, R.L., 1996. Quadrupole Mass Spectrometers: An Intuitive Look at the Math. Journal of Chemical Education, 73: 1142.

Macek, B., Waanders, L.F., Olsen, J.V. and Mann, M., 2006. Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. Mol Cell Proteomics, 5: 949-58.

March, R.E., 1997. An Introduction to Quadrupole Ion Trap Mass Spectrometry. Journal of Mass Spectrometry, 32: 351-369.

Marina, A. et al., 1999. High-sensitivity analysis and sequencing of peptides and proteins by quadrupole ion trap mass spectrometry. J Mass Spectrom, 34: 17-27.

Marrie, T.J., Kwan, C., Noble, M.A., West, A. and Duffield, L., 1982. Staphylococcus saprophyticus as a cause of urinary tract infections. J Clin Microbiol, 16: 427-31.

Mateo, M. et al., 2005. Genotypic versus phenotypic characterization, with respect to susceptibility and identification, of 17 clinical isolates of Staphylococcus lugdunensis. J Antimicrob Chemother, 56: 287-91.

Mendoza, M., Meugnier, H., Bes, M., Etienne, J. and Freney, J., 1998. Identification of Staphylococcus species by 16S-23S rDNA intergenic spacer PCR analysis. Int J Syst Bacteriol, 48 : 1049-55.

Morgan, M.C., Boyette, M., Goforth, C., Sperry, K.V. and Greene, S.R., 2009. Comparison of the Biolog OmniLog Identification System and 16S ribosomal RNA gene sequencing for accuracy in identification of atypical bacteria of clinical origin. J Microbiol Methods, 79: 336-43.

Noble, W.C., 1984. Skin Microbiology: Coming of Age. Journal of Medical Micobiology, 17: 1-12.

Noguchi, N. et al., 2010. Using the tannase gene to rapidly and simply identify Staphylococcus lugdunensis. Diagn Microbiol Infect Dis, 66: 120-3.

Ohara-Nemoto, Y., Haraga, H., Kimura, S. and Nemoto, T.K., 2008. Occurrence of staphylococci in the oral cavities of healthy adults and nasal oral trafficking of the bacteria. J Med Microbiol, 57: 95-9.

Pereira, E.M., Oliveira, F.L., Schuenck, R.P., Zoletti, G.O. and Dos Santos, K.R., 2010. Detection of Staphylococcus lugdunensis by a new species-specific PCR based on the fbl gene. FEMS Immunol Med Microbiol, 58: 295-8.

Phillips, C.I. and Bogyo, M., 2005. Proteomics meets microbiology: technical advances in the global mapping of protein expression and function. Cell Microbiol, 7: 1061-1076.

Poyart, C., Quesne, G., Boumaila, C. and Trieu-Cuot, P., 2001. Rapid and accurate species-level identification of coagulase-negative staphylococci by using the sodA gene as a target. J Clin Microbiol, 39: 4296-301.

Rhoden, D.L. and Miller, J.M., 1995. Four-year prospective study of STAPH-IDENT system and conventional method for reference identification of Staphylococcus, Stomatococcus, and Micrococcus spp. J Clin Microbiol, 33: 96-8.

Roth, R.R. and James, W.D., 1988. Microbial ecology of the skin. Annu Rev Microbiol, 42: 441-64.

Satta, G., D'Andrea, L., Grazi, G., Soro, O. and Varaldo, P.E., 1993. Micrococci demonstrate a phosphatase activity which is repressed by phosphates and which can be differentiated from that of staphylococci. Int J Syst Bacteriol, 43: 813-8.

Scherer, R.K. and Brown, R.W., 1974. Differentiation of staphylococcal and micrococcal proteinases by electrophoresis. Appl Microbiol, 28: 768-73.

Schumann, P., Kampfer, P., Busse, H.J., Evtushenko, L.I. and Subcommittee on the Taxonomy of the Suborder Micrococcineae of the International Committee on Systematics of, P., 2009. Proposed minimal standards for describing new genera and species of the suborder Micrococcineae. Int J Syst Evol Microbiol, 59: 1823-49.

Shah, N.B. et al., 2010. Laboratory and clinical characteristics of Staphylococcus lugdunensis prosthetic joint infections. J Clin Microbiol, 48: 1600-3.

Shevchenko, A. et al., 1996. A strategy for identifying gel-separated proteins in sequence databases by MS alone. Biochem Soc Trans, 24: 893-6.

Sivadon, V. et al., 2005. Use of genotypic identification by sodA sequencing in a prospective study to examine the distribution of coagulase-negative Staphylococcus species among strains recovered during septic orthopedic surgery and evaluate their significance. J Clin Microbiol, 43: 2952-4.

Skoog, D.H., FJ. Crouch, SR, 2007. Principles of Instrumental Analysis. Thomson Learning.

Skow, A. et al., 2005. Species-level identification of staphylococcal isolates by real-time PCR and melt curve analysis. J Clin Microbiol, 43: 2876-80.

Spanu, T. et al., 2003. Use of the VITEK 2 system for rapid identification of clinical isolates of Staphylococci from bloodstream infections. J Clin Microbiol, 41: 4259-63.

Stackebrant, E., Cathrin Kock, et. al., 1995. Taxonomic Dissection of the Genus Micrococcus: Kocuria gen.

nov., Nesterenkonia gen. nov., Kytococcus gen. nov., Dermacoccus gen. nov., and Micrococcus Cohn 1872 gen. emend. International journal of Systemic Bacteriology, 45: 682-692.

Stafford Jr, G.C., Kelley, P.E., Syka, J.E.P., Reynolds, W.E. and Todd, J.F.J., 1984. Recent improvements in and analytical applications of advanced ion trap technology. International Journal of Mass Spectrometry and Ion Processes, 60: 85-98.

Suckau, D. et al., 2003. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Analytical and Bioanalytical Chemistry, 376: 952-965.

Surekha.Y.Asangi , M.J., Sathyanarayan.M.S , Nagabhushan , Rashmi, 2011. Speciation of clinically significant Coagulase Negative Staphylococci and their antibiotic resistant patterns in a tertiary care hospital. Int J Biol Med Res., 2: 735-739.

Tanaka, K. et al., 1988. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry, 2: 151-153.

Tang, J.S., 2003. Reclassification of ATCC 9341 from Micrococcus luteus to Kocuria rhizophila. International Journal of Systematic and Evolutionary Microbiology, 53: 995-997.

Van Riper, S.K., de Jong, E.P., Carlis, J.V. and Griffin, T.J., 2013. Mass spectrometry-based proteomics: basic principles and emerging technologies and directions. Adv Exp Med Biol, 990: 1-35.

Whitehouse, C.M., Dreyer, R.N., Yamashita, M. and Fenn, J.B., 1985. Electrospray interface for liquid chromatographs and mass spectrometers. Anal Chem, 57: 675-9.

Wieser, M. et al., 2002. Emended descriptions of the genus Micrococcus, Micrococcus luteus (Cohn 1872) and Micrococcus lylae (Kloos et al. 1974). Int J Syst Evol Microbiol, 52: 629-37.

Wollnik, H., 1993. Time-of-flight mass analyzers. Mass Spectrometry Reviews, 12(2): 89-114.

Yates, J.R., 3rd, 1998. Mass spectrometry and the age of the proteome. J Mass Spectrom, 33: 1-19.

Yugueros, J. et al., 2000. Glyceraldehyde-3-phosphate dehydrogenase-encoding gene as a useful taxonomic tool for Staphylococcus spp. J Clin Microbiol, 38: 4351-5.