8-2020

Metabolomics Advanced and Applied: Shrinking the metabolomic knowledge gap by advancing and applying metabolomics in integrated multi-omics

Matthew J. Keller
University of Tennessee

To the Graduate Council:

I am submitting herewith a thesis written by Matthew J. Keller entitled "Metabolomics Advanced and Applied: Shrinking the metabolomic knowledge gap by advancing and applying metabolomics in integrated multi-omics." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Chemistry.

Shawn R. Campagna, Major Professor

We have read this thesis and recommend its acceptance:

Christopher A. Baker, Thahn D. Do, Robert L. Hettich

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Metabolomics Advanced and Applied: Shrinking the metabolomic knowledge gap by advancing and applying metabolomics in integrated multi-omics

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Matthew Joseph Keller

August 2020

Acknowledgements

I want to thank Dr. Shawn Campagna for serving as my advisor. You took me into your lab and have been supportive of my research and studies. I have learned a lot in the time I have spent in your lab and am very thankful for the opportunities you gave me. I would also like to thank Drs. Chris Baker, Thahn Do, and Robert Hettich for serving on my committee. I would like to thank Drs. Stephen Kania, Mohamed Abouelkhair, and Robert Hettich as well as David Reeves for their collaborations in my research. I additionally am thankful for Dr. Hettich providing my funding in this last semester.

I wish to thank Dr. Hector Castro for his guidance and keeping operations in the BSMMSC running smoothly. I know that you truly care for us grad students and I appreciate you looking out for me during my tenure in UT Chemistry. I want to thank all of the Campagna group members I have worked with, including Dr. Eric Tague, Dr. Caleb Gibson, Dr. Brandon Kennedy, Dr. Alexander Fisch, Dr. Joshua Powers, Ashley Lato, Katarina Jones, Katharina Höland, Courtney Christopher, Zarin Tasnim, Brittni Woodall, Lindsay Brown, Zane Vickery, and Westley Seaton. I am especially grateful for Drs. Tague and Powers for teaching me a lot about analytical chemistry and mass spectrometry. I would also like to thank my undergraduate advisor Dr. Jisook Kim for her guidance in my undergraduate education, and my undergraduate PI Dr. John Lynch for getting me started on research.

I am extremely thankful for my family who has supported me for the years. Without my parents, Calvin and Connie, I doubt very much that I would have ever gone to graduate school, let alone complete a degree. I need to thank my wife, Rose, who has been such a wonderful blessing to me. I am deeply thankful for the love and grace you have shown to me.

Finally, I must thank my Lord and Savior Jesus Christ for his amazing mercies to me his sinful child. He has orchestrated all these blessings that I have enjoyed through my life and by his blood I have the confidence of eternal life with him.

Abstract

Some of the most exciting questions in chemistry lay within the realm of molecular biology. Although different disciplines, throughout history we see chemists and chemical techniques leading the way in important biological discoveries. Metabolomics is a new, developing technique in molecular biology that is spurred on by technical innovations, primarily from the chemistry and engineering fields. Here, two different liquid chromatography mass spectrometry techniques and state-of-the-art bioinformatic tools are employed to help expand the field of metabolomics. In the application phase of this project, metabolomic techniques were applied in a multiomic experiment to elucidate the metabolic pathways used in *Staphylococcus*. Multi-omics are the coupling of multiple omics techniques such as metabolomics, genomics, and proteomics. In particular, an Ultra-Performance Liquid Chromatography-High Resolution Mass Spectrometry platform was used with a semi-targeted metabolomics technique. Large fold-changes are observed in metabolites mevalonate and phosphomevalonate, which are important distinguishing metabolites between the two isoprenoid synthesis routes. This is used to characterize isolates based on which metabolic pathway they use. This is further verified and expanded by the use of comparative genomics. In the developmental phase of this project, metabolomics techniques were advanced by testing and comparing different extraction methods for multiomic analyses. In this case, chloroform-based extractions were tested against methyl-*tert*-butyl ether-based extractions to collect metabolites, lipids, and proteins simultaneously. This was analyzed using a High-Performance Liquid Chromatography-High Resolution Mass Spectrometry platform with split-flow nano chromatography and electrospray ionization. Additionally, the cell lysis method is investigated to determine its impact on extraction efficiency and metabolite degradation. Optimizing extraction procedures will make multi-omics faster, easier, and more reliable, thereby facilitating greater use of metabolomics in multi-omics experiments.

Table of Contents

List of Tables

List of Figures

Chapter 1: A Brief Introduction to Liquid Chromatography Mass Spectrometry-Based Metabolomics

The Task

Genomics

What is metabolomics? What questions does it seek to answer? Why would a bioanalytical chemist be interested in it? To answer these questions, we must first start at a surprising place, the Human Genome Project (HGP, https://www.genome.gov/human-genome-project). The genomics revolution was driven by technological advancements that allowed genomes to be rapidly and accurately sequenced. With the completion of the HGP and the sequencing of the human genome, many people thought vast numbers of diseases would be explained and rapidly cured. Despite expectations, however, there have been relatively few improvements to modern medicine. Those who had high hopes for the HGP underestimated several factors.

First, despite dramatic improvements and the maturation of genomics as a field, there are still technical hurtles that can be challenging to genomics studies. Larger eukaryotic genomes are known to have large stretches of highly repetitive sequences in chromosomes such as satellites. These can make full sequencing problematic, as it is difficult to stitch together contigs when the separating sequences are extremely large. Indeed, if there is too much repetition of any kind (such as polyploidy common in plants), the technical and informatic burden become great, and sequences are often prohibitively difficult to assemble.

The second limitation of genomics is that it is not yet possible to unambiguously explain gene function from sequence alone, at least not *ab initio*. One difficulty in explaining function is that the DNA sequence does not directly relate to amino acid sequence. This is the case because after transcription to mRNA, introns (non-amino acid coding sections of the gene) are removed and the exons (coding sections) are stitched back together. This changes the final sequence from what one would expect by analyzing the gene sequence alone. Finally, protein folding and post-translational modifications also must be taken into account which add a further confounding layer of complexity. When it comes to protein biochemistry, structure is critical to determine function. Without knowing how the polypeptide will fold and wrap together, it is challenging to elucidate the function from the sequence alone.

Thirdly, there are a vast number of different regulatory mechanisms governing post-translation modifications (PTM), as well as which genes are active. Transcriptional regulation is a complex process that determines which genes are transcribed into mRNA to code for proteins. Many genes are only activated in response to certain stimuli, and from the genetic sequence, it is impossible to know which genes are active when. Furthermore, after being translated, many modifications can occur. For example, for collagen to form its functional trimeric form, many PTMs must occur, including PTMs like glycosylation and hydroxylation (a process involving at least nine enzymes).[1] The failure of this process leads to the disease known as scurvy. This is not an isolated event, many different PTMs occur regularly in biological systems including: phosphorylation, methylation, glycosylation, and acylation, among others.[2] This complexity again separates protein function from genetic code, making a complete and dynamic biological interpretation impossible.

Finally, a simplistic genomics view of physiology fails to account for the role of environmental factors feeding into this network of regulatory and signaling pathways. This can occur at many levels through epigenetics, hormone signaling, and metabolic inputs. One must dig deeper to get an understanding of the biological state of an organism. Instead of merely sequencing the genome, in order to get more direct biological information, many attempt to interrogate which proteins exist at what concentrations. This is the field of proteomics.

Proteomics

Proteomics involves the global analysis of all proteins in a cell, tissue, organ, or even entire organism. In its origins, proteomics was based on 2-D gel separation techniques and measured proteins with stains and antibodies.[3] This allowed for only rudimentary characterization of proteomes. However, fueled primarily by innovations in instrumentation, proteomics has developed into a mature field using two-dimensional high performance liquid chromatography (LC) coupled to high-resolution tandem mass spectrometers (MS).[4-7] In certain systems (like yeast), a nearly complete proteome coverage can be achieved with these methods.[8] Measuring relative levels of these proteins when comparing an experimental group to a control group can be very powerful in elucidating impacts of the treatment on the experimental group. This is commonly done to elucidate gene function (proteogenomics), probe for biomarkers, and

in pharmaceutical research.[9-11] While this approach more closely probes molecular function, proteomics also has limitations.

Despite the development of robust methodologies, there are still technical difficulties for proteomics. One such problem is that of quantification. Commonly, quantification is only performed in a relative manner, comparing two or more test conditions. Additionally, these values are based on peak areas of the MS^1 data for peptides (in bottom-up, the most common proteomic technique). However, this can be misleading because of isobaric coeluting species that can greatly impact the accuracy of this quantification. Additionally, full-scans need to be balanced with fragmentations to both provide both deep peptide sequences and give accurate quantitative values.[12] Additionally, there is no good way to go from peptide quantifications to protein quantification for large numbers of proteins due to differential ionizations, enzymatic cleavage efficiency, and protein/peptide losses across experimental steps. There are other challenges as well such as peptide identification, adequate sampling of membrane proteins, and accurately preserving post-translation modifications, but space does not permit a full description.

Beyond the experimental and informatic challenges of proteomics, many proteins detected have unknown functions and kinetics. Additionally, it is difficult to predict how the thousands of proteins will interact to form the metabolic state of an organism. Experiments bear this out, where predicted states based on proteomics or transcriptomic data conflict with actual studies of the metabolites.[13-14] This is hypothesized to be due to metabolic regulation being controlled through enzymatic efficiency, not capacity (pool-size).[13] This is the gap metabolomics desires to fill—to discover gene functions, to characterize proteins, to elucidate signaling pathways, and determine the physiological impact of small molecules. Although there are different methods of attempting this analysis (nuclear magnetic resonance, Fourier transform infrared spectroscopy, etc…) we will be constraining our focus to liquid chromatography-mass spectrometry-based techniques. These are the most commonly used and one of the most information dense.

The Challenges

Metabolomics in its fullest sense is the global analysis of all small molecules in a biological system (metabolites).[15] This is an astoundingly ambitious goal. Thousands of metabolites are already known with the Human Metabolome Database listing 9037 metabolites

that have been detected in biological specimens (as of 1/9/2020).[16] However, myriads more remain unknown and unverified with estimates of unique metabolites as high as 200,000.[15] Identifying and characterizing these compounds would open up worlds of biological knowledge; however, this presents several challenges. The first is that the possibilities are virtually endless. In the field of proteomics, peptide identification is greatly simplified by the presence of the genome. Proteins come directly from the genes (albeit some display modifications) and so by sequencing the genome, a reference guide of protein possibilities is generated. However, there is no such reference when it comes to metabolomics. Genetic or protein sequences do not provide a direct guide as to what reactions (and thereby what molecules) occur. Such determinations must be done through painstaking experimentation. Even if mass spectrometric measurements could be done to such a degree of accuracy that unique formulae were assigned to each spectral feature (which would require mass accuracies well below 0.1 ppm[17]—not possible outside of specialized ion cyclotron resonance instruments for the foreseeable future), the analyst is still far from a chemical structure, much less stereochemistry. Additionally, the lack of a "reference manual" means metabolite origin is difficult to determine. Metabolites analyzed from bovine serum for example, may be produced from bovine tissues or from the microbiota of its rumen. The same is true for plant-microbe interactions and other symbiotic organisms. Unclear origins such as these can be important depending on the biological question being asked.

Not only are there vast numbers of compounds, with seemingly endless possibilities, but these span a broad range of polarity and other chemical properties. From glucose to cholesterol to odd compounds like cyanocobalamin (vitamin B12), metabolites are exceedingly diverse. Many of these metabolites are labile or volatile, which further complicates their capture and analysis. Beyond chemical variance, metabolites span a wide concentration range with estimates of at least twelve orders of magnitude (the best mass spectrometers can only touch six orders of magnitude).[18]

The sheer number of metabolites also provides obstacles to analysis. Inevitably, metabolites coelute, which modern time of flight (TOF) and orbitrap instruments are able to resolve (except for compounds with the same mass), but if tandem MS data are desired, instrument duty cycles are not always fast enough to capture all coeluting compounds. Similarly, most instruments are not able to switch polarities mid-run effectively, which hampers broad metabolite coverage as some molecules feasibly ionize only in negative mode while others only

ionize in positive mode. This requires multiple analyses to fully capture the diversity of small molecules.

Prognosis

To summarize the current state of metabolomics, both the total number and identify of most metabolites is unclear; there is no single method capable of extracting all metabolites from a system, let alone analyze them; mass spectrometers cannot uniquely identify new metabolites or span the concentration ranges metabolites display and they lack the duty cycle to comprehensively annotate all metabolites in a mixture; finally, even with measured spectral features, we lack the ability to easily distinguish true signals from chemical and instrumental noise or to determine the biological significance of those signals. In short, a metabolomics experiment is doomed from the start.

Although the ideal metabolomics experiment is far out of reach, we can divide this lofty goal into more manageable pieces. One customary division is the separation of lipids from polar metabolites. Although lipids are rightly considered a subsection of metabolites, they are convenient to separate from conventional water-soluble metabolites because of polarity differences. Lipid are relatively non-polar compounds (though there is no strict definition, they are generally agreed to be amphiphilic or hydrophobic hydrocarbon-based molecules). For this reason, aqueous-based extractions will miss much of the lipidome (an organism's complete set of lipids), and it is typically better to dedicate a separate method for lipid analysis rather than hamstring the metabolomic analysis by trying to include lipids. Hydrophilic metabolites are often measured using various LC-MS based techniques (although capillary electrophoresis is sometimes used).[19-21] LC is a technique that separates analytes based on differential affinities as they pass through a column. The eluent is then directed to a mass spectrometer (usually using electrospray ionization) that measures the compounds. The mass spectrometers measure mass to charge ratio (*m/z*) and assign intensities to each *m/z*. This is sometimes done in tandem where the molecule is fragmented, and the fragments measured. Metabolites are challenging to characterize for the reasons described above, including their chemical diversity, concentration ranges, and the lack of available information. Because of this, studies will often employ more than one analytical technique, combining different chromatography and mass spec polarities to increase metabolite coverage.

For lipid analyses, the lipidome is often further subdivided with amphiphilic, or charged, lipids (phospholipids for example) and hydrophobic, or neutral, lipids (steroids being a major class), again for practical reasons. Although often thought of as performing merely structural or energy roles in physiology, lipids are being found more and more to also play important signaling and regulatory roles. This is already well known in the case of steroid hormones, but others have been characterized including sphingolipids, eicosanoids, and inositol phospholipids.[22-24] Amphiphilic lipids are generally analyzed by RPLC-MS, HILIC-MS, and direct infusion MS methods.[25-28] Nonpolar lipids, on the other hand, have been analyzed by nonaqueous RPLC as well as normal phase methods.[29-31]

Another practical subdivision of metabolomics is to dedicate specific methods to volatile metabolites. Many metabolomic extraction procedures involve a drying step in which volatiles are lost. These are best analyzed by gas chromatography (GC)-MS based methods. GC is a gas-phase separation that is naturally suited to volatiles. However, due to the gas phase requirements and the high temperature, this is unsuited to non-volatile or thermally labile metabolites (such as sugars or nucleotides). It is worth noting, however, that non-volatile metabolites and lipids are sometimes analyzed by GC-MS using derivatization techniques, however, this introduction focuses on LC-MS methods.

Beyond methods dedicated to certain classes of molecules, the analysis can be further focused by separating “targeted” and “non-targeted” analysis. In targeted lipid/metabolomics, the analysis is limited to a set of known compounds where absolute quantitative data are desired (where a value in molarity or g/L is generated). Non-targeted or discovery metabolomics instead tries to analyze an extensive array of metabolites, including unknown metabolites, while only employing relative quantification (relative quantification is fold change relative to a control). Generally, to get highly accurate reproducible quantitative data, the method must be set up to optimize (target) the instrument toward measuring the compounds of interest. If this is done, a global analysis is hampered by the method bias and cannot be as comprehensive. The method employed is dependent upon the question being asked. Often non-targeted metabolomics are used for “hypothesis generating” experiments. Targeted metabolomics, on the other hand, are often used to test hypotheses. Thus, it is not unusual for a targeted metabolomics experiment to be performed to explore and validate findings from a non-targeted experiment.

Important Developments

There has been headway made in the pursuit of these sub-goals. Oliver Fiehn, along with Tobias Kind and others in his lab have established principals to guide unknown spectral feature formula assignment, facilitating untargeted analyses. They were the first to demonstrate that mass accuracy will never (in the foreseeable future) give the specificity required for unique formula assignment.[17] Following up on that, in their "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry"[32] paper, they describe seven rules which are helpful to reduce the number of possibilities generated from matching chemical formulae to experimental masses. The Fiehn lab has also established a popular metabolomics database containing chromatographic (GC) and tandem mass spec information for a wide variety of metabolites, the FiehnLib.[33]

Regarding the separation half of the analysis, Joshua Rabinowitz and his lab made a significant contribution by their development and validation of chromatographic methods for metabolomics. In particular, he developed both HILIC-MS and RP-MS methods for non-targeted metabolomics.[34-35] The RP method employs an ion-pairing reagent to improve retention of polar metabolites on the nonpolar column but keeping the robustness of a reversed phase packing.[35] This method has been utilized by others, demonstrating the versatility of the method.[36-40]

One of the greatest achievements of the lipidomics field is the development of shotgun lipidomics. Unlike most methods utilizing separations, this involves the direct infusion of lipid-rich extracts into the mass spectrometer. As such, this method is faster, easier, and much less expensive than those which employ chromatography (no solvent or apparatus costs). This was first developed by Han and Gross in 1994.[27] They identified greater than 50 lipid species from human erythrocyte membranes with quantitative accuracies >95%. This approach is still used today for rapid profiling of phospholipid species.

Overall, there have been strong developments made in metabolomics as a field, although there is still much to be done. We will now consider these advances and the current state of the field in more detail.

Experimental Approaches

Metabolomic Extractions

To begin any metabolomics analysis, the first step must be extraction. To analyze metabolites, they must first be freed from the protective shell of the cell wall/membrane. As discussed, this looks different depending on the type of analysis being performed. Polar metabolites represent the most diverse and problematic of the metabolomic subsets. Several challenges immediately present themselves. First, in order to accurately represent the intracellular components of the sample (be it a culture, tissue, or other biomass) in the extract, the procedure must quench metabolism (stop enzymatic activity). To quench metabolism, temperature shocks are often employed, using liquid nitrogen or chilled solvent, but flash freezing is required if there is a delay between sample harvesting and extraction.[41-43] Additionally, the extraction method must be careful not to restore enzymatic activity later in the method. This is particularly an issue when the polar metabolites are concerned because aqueous-based solvents do not halt enzymatic activity and the metabolic state will be perturbed. Therefore, most extraction solvents employ a polar organic component in the solvent to denature and precipitate proteins. Sometimes high or low pH is used to accomplish the same effect. However, the use of pH extremes for protein precipitation has been reported to result in lower reproducibly measured features, possibly due to degradation, co-precipitation, or ionization suppression of metabolites, or a combination thereof.[44-47]

The extraction method and solvents also need to lyse cells and limit abiotic degradation. Solvents used for decades include perchloric acid,[48] hot water,[49] and boiling ethanol[50] which were originally used to measure intracellular amino acid content. These are relatively harsh methods and since 1990 softer extractions method have been developed to prohibit metabolite degradation.[51] Many studies employ some kind of chloroform/methanol extraction (CME). This is a biphasic extraction that can be utilized for collection of either metabolites or lipids. Cold methanol, chloroform, and water are sequentially added to the sample and mixed for several minutes to permeabilize and extract the contents of the cell.[42] This methanol/chloroform mixture has the added benefit of precipitating proteins to stop potential enzymatic activity. After phase separation and centrifugation, the aqueous layer is collected for metabolite analysis.

Perhaps due to the inconvenience of this lengthy extraction, or to avoid the use of chloroform, a probable human carcinogen, a cold methanol/water extraction protocol was

developed.[52] This uses a simple 50/50 methanol/water solvent and freeze-thaw cycles to lyse cells and became commonly used.[53] However, there have been some concerning reports questioning the efficacy of this extraction method. Rabinowitz and Kimball found that the cold methanol procedure "resulted in marked decomposition of nucleotide triphosphates."[53] Canelas et al. also found that cold methanol with freeze-thaw cycles did not stop enzymatic activity.[51] This is perhaps unsurprising since enzymatic activity has been assayed down to −100 °C in methanol solutions and may explain the observed loss of nucleotide triphosphates.[51, 54] Rabinowitz and Kimball developed an alternate extraction method to preserve unstable compounds like the triphosphates.[53] This employed an acidic (0.1 M formic acid) acetonitrile-based extraction (AAE) with either water or methanol and water as additional components in the solvent. They found that AAE performed better than cold methanol by limiting degradation and provided better yields for almost all metabolites than did CME.[53] However, Canelas et al. tested AAE against CME and found opposite results.[51] They attribute this discrepancy to the lack of mixing when Rabinowitz and Kimball tested CME as well as differences between model organisms (Rabinowitz used *E. coli* and Canelas used *S. cerevisiae*). In CME, the chloroform, methanol, and water will phase-separate into two layers requiring a period of mixing to achieve equilibrium partitioning of analytes. Without this mixing, it is conceivable that the extraction efficiency would decrease.

CME preparations were first used for lipid analyses. Made popular by Folch's papers in 1951 and 1956, CME variants are still very common for lipid extractions.[55-56] Folch's method took advantage of the insolubility of chloroform in water to develop a two-phasic extraction similar to those already discussed that washes out impurities and leaves lipids in the chloroform layer. It used relatively large solvent amounts in an 8:4:3 ratio of chloroform, methanol and water. Bligh and Dyer developed a variant of the Folch method which is rapid and uses less solvent than Folch's method.[57] The general method is similar and uses a final mixture of 2:1:1 chloroform, methanol, and water to extract lipids and wash contaminates. This seems to be very similar to the Folch method and may work better when the sample has a low mass fraction of lipids. At high mass fractions, however, Bligh and Dyer's method does not give as high yields as Folch's which makes sense since the smaller volume solutions may be saturating.[58]

The diversity of extraction methods used complicates the comparison of different metabolomic studies. Are different findings the result of actual differences in the systems being

studied or are they artifacts of differences in the procedures used? Tentatively, CME seems to be the most reliable extraction method for metabolites when done correctly, although other methods may be better if a faster analysis is desired. Whereas for lipids, the Folch or Bligh and Dyer methods both work, depending on the sample. More work still needs to be done to validate and standardize extraction methods. In the meantime, however, data can be reasonably compared to each other for results generated in the same experiment. Although the absolute levels may be skewed from the original biological state (ATP/ADP ratio, for example), differences between samples ought to be reflective of real differences in the organism. However, particularly for metabolites known to be unstable, observed differences may be secondary, and so one must be careful when inferring biological conclusions.

Recently, there has been increasing interest in developing extraction methods to simultaneously collect multiple classes of biological molecules. This is applicable to "muli-omic" analyses where multiple omics experiments are performed in the same study. For example, a recent study employed metabolomics, lipidomics, and proteomics to elucidate the functions of mitochondrial genes/proteins.[59] The benefit of doing a concerted extraction is that each molecular biome derives from the same biological state. Otherwise, it is hard to ensure different extractions are harvested at identical physiological conditions. Performing one extraction to collect three (or more) molecular classes also greatly reduces sample consumption and speeds sample collection. This is particularly advantageous when sample volumes are at a premium (as is often the case with clinical samples). The disadvantage is that the procedure cannot be optimized to focus on any particular set of compounds. This may reduce extraction efficiencies when compared to focused extraction procedures. Two such tandem extraction protocols were published in 2016, the so-called SIMPLEX (Simultaneous Metabolite Protein Lipid Extraction) and MPLEx (Metabolite, Protein, and Lipid Extraction) methods.[60-61] These are both modifications of previously existing extraction methods, with SIMPLEX based on the Matyash lipid extraction (a biphasic MTBE/methanol/water extraction) and MPLEx based on the CME.[56, 62] With SIMPLEX, the resulting extraction has lipids in the organic top layer, metabolites in the aqueous lower layer, and proteins pelleted at the bottom of the vial. When MPLEx is used, due to chloroform's extreme density, the organic layer sits at the bottom of the vial and aqueous layer is at the top with proteins sandwiched between them. SIMPLEX expedites the process slightly by making fraction collection easier by pelleting out the proteins. However,

there has been no direct comparison of these two methods, and it is unclear what the different efficacies might be.

Fractionation

Extraction yields a very complex mixture of thousands of chemicals. For this to be analyzed precisely, the sample must be fractionated. A fractionation step helps to reduce ion suppression (an effect where certain molecules preferentially ionize over others, drowning their signal) and wash out other interfering species like salts, which may be present in the extract. If tandem MS data is desired, fractionation also allows time for analytes to be sequentially selected for MS^2 analysis. There are drawbacks to a pre-MS fractionation, such as sample dilution and increased analysis time. However, the majority of the time, the benefits outweigh the negatives. In early metabolomics work (called metabolic fingerprinting), gas chromatographs were often employed.[63-66] GC continues to be employed for metabolomic studies today, although LC-based methods are now more common.[47, 67-69] Much could be said regarding GC metabolomics, but here we will be constraining our focus to LC-MS metabolomics. It is worth noting that GC metabolomics requires derivatization (for broad metabolite coverage) and still cannot be used for large or labile compounds.[15] LC has also been used for metabolomic analysis. LC-MS started to be applied to metabolite analyses in earnest at the turn of the millennia (early 2000s or just before). These were focused studies, but such work laid the foundation for further expansion.[70-73] Reversed phase (RP) separations were commonly used due to their ease-of-use and column robustness. Retention presents a problem, however. RP separates based on analyte adsorption to a hydrophobic stationary phase, which can work for lipids (being relatively non-polar), but metabolite methods often suffer from poor retention. This leads to coelution of metabolites and elution in the dead-volume. Having several coeluting compounds increases the instrumental burden as separation does not resolve compounds with similar or identical masses. Second, salts and other contaminates often elute in the dead-volume, which causes ion suppression for early eluting analytes. Lu et al. report a method to quantify 90 metabolites using a Fusion-RP column (Phenomenex) with polar elements incorporated into the C18 stationary phase.[74] The embedded polar elements provide some selectivity to polar metabolites allowing separation; however, even this is still hampered by retention issues and poor peak shape.[75] Further improvements, were needed and there are two main substitutes to achieve superior separations for polar metabolites.

The first is Hydrophilic-Interaction Liquid Chromatography (HILIC). This uses a polar stationary phase (often silica), which builds up a water rich solvent layer.[76] The dominant separation mechanism was thought to be through differential partitioning between the water layer (mostly static) and the organic rich mobile phase.[77] However, other mechanisms are at play as well, including adsorption and ion exchange.[77] HILIC gives superior retention for metabolites compared to RP, but chromatographic peak profiles for HILIC separations can often be poor. Bajad et al. in 2006 found good separation and peak shapes with an amino HILIC column.[34] The amino functional groups on the stationary phase used are reactive, however, and the column lifetime is short. The second substitute takes the concept of adapting RP chromatography to polar analytes further. Lou et al. (later adapted by Lu et al.) added an ion-pairing reagent to the chromatography solvents (tributylamine, in this case).[35, 78] This increases retention by forming a complex with charged analytes and the nonpolar functionality (butyl groups) serve to increase affinity to the stationary phase. By changing the selectivity this way, most metabolites elute past the dead volume and are more dispersed across the gradient. However, ion pairing reagents are not ideal for MS use. This is because they are permanently charged species that will cause severe ion suppression in their incompatible instrument polarity (positive mode for tributylamine). Additionally, ion pairing reagents tend to be extremely recalcitrant and stick to LC fittings, columns and MS components. The means that the ion suppression will persist even when using separate solvents. Despite these handicaps, some labs feel the benefits outweigh the complications and Lu's ion pairing method has been used for wide metabolome coverage by the Campagna lab and others.[38, 79-80]

For lipid fractionation, the methods used will vary based on the specific analytes of interest. For phospholipids (the most common amphiphilic lipids), fractionation is not always used as discussed in shotgun lipidomics. However, to improve coverage of amphiphilic lipids or simplify analysis chromatography is employed. RP methods have been developed for phospholipids. These roughly separate lipids based on tail chain length. However, based on the hydrophobic interactions between the fatty tails and aliphatic stationary phase, it can be difficult to elute lipids as they bind strongly to RP stationary phases. HILIC is also employed, which roughly separates lipids based on headgroup polarity. These interactions are less strong than with RP, allowing lipids to be more easily eluted from the column. This selectivity allows sequential analysis of lipid classes, simplifying analysis, especially when fragmentation is not employed.

The hydrophobic lipids are more challenging. RP methods have been attempted; however, many adaptations are required to avoid irreversible binding. Cai et al. developed a non-aqueous RP technique, which uses APPI (atmospheric pressure photo ionization—one of the better methods of ionizing hydrocarbons) for MS analysis.[29] They were working with triacylglycerols (one type of neutral, or nonpolar, lipids) and found that many nonaqueous solvent systems worked well but recommended acetonitrile/isopropyl alcohol for practical reasons. APPI with solvent additives (acetone) was demonstrated to give good signal with estimated LODs below 200pg. Normal phase chromatography-based methods have also been employed with good results. Hutchins et al. report a method for separation and analysis of cholesteryl esters, triacylglycerol, diacylglycerols, and monoalkyl-ether diacylglycerols using an MTBE/hexane gradient with a silica column.[31] This was coupled to a MS by use of a splitting T and addition of an electrospray modifier solution. This was necessary because aprotic-nonpolar solvents such as hexane and MTBE do not spray in ESI.

Mass Spectrometry

Just as genomics was driven by technological and methodological innovations, so too metabolomics depends on development of analytical platforms sophisticated enough to handle the complex samples. Early metabolomics was done with triple quadrupole (QQQ) mass spectrometers. A quadrupole is a mass analyzer that filters out ions except for a particular *m/z*. If the ion flux is detected, an intensity can be assigned to the selected *m/z*. In a QQQ, the first quadrupole is used to select an ion (precursor or parent ion), which is fragmented in a second quadrupole, or higher order multipole, (which does not filter ions) and a product (or daughter) ion is selected in the third quadrupole. Depending on how the quadrupoles are operated, various different MS data can be generated with common modes including selected reaction monitoring (where specific parent and product ions are preset), product ion scan (where all possible product ions are scanned for particular parent ions), and neutral loss scans (where parent and product ions are scanned with a set *m/z* offset to see what ions lose a particular group). Due to the quadrupole's low resolving power and cumbersome operation with large numbers of analytes, non-targeted profiling is functionally impossible and QQQ metabolomics are often operated with selected reaction monitoring of a few dozen targeted metabolites. Technological innovations have exploded analytical capabilities with the introduction of high resolution TOF and orbitrap

instruments. TOF instruments analyze ions by pushing them down a flight tube and measuring how long it takes to reach the detector. The signal intensity of the detector relates to the number of ions and the time relates to the *m/z*. When optimized, TOFs can have high resolution capabilities. An orbitrap injects packets of ions into a small cell with an electrode in the center. Ions orbit the electrode and oscillate along the length of the cell due to the shape of the cell. This oscillation is *m/z* dependent and analysis can be performed by measuring the induced current (image current) as ions oscillate across the football-shaped ion cell. This signal can be deconvoluted by a Fourier transform to generate ultra-high resolution data. These also have full scan capabilities, where the full mass spectrum (within the operating parameters of the instrument) is measured every scan. The resolving power improvements combined with full scans gives incredibly richer data sets than with quadrupoles. This allows for analyses of hundreds and potentially even thousands of compounds in the same experiment, even detecting unknown compounds, thereby unlocking the possibility of truly non-targeted experiments. Hybrid variants such as the Q-TOF and Q-orbitrap (the QE, or Quadrupole-Exactive, instrument) further improve analyses and are used more often for metabolomics than their single stage versions. These also allow for tandem MS to be performed except that the MS^2 are performed in multiple reaction monitoring (MRM—where all possible daughter ions are detected) due to the full-scan capabilities of these instruments. For non-targeted experiments, these will often be operated in a data-dependent acquisition (DDA) mode. With DDA, the n-most intense ions of a full scan are sequentially selected for fragmentation (MRM in this case). This allows for profiling of unknowns as features can be selected for fragmentation without being known *a priori*. To efficiently perform DDA, a fast duty cycle is required such that a large number of ions can be selected for fragmentation in a short amount of time. Otherwise, rapidly eluting molecules might be missed by the MS^2 scans. Or, even if they are measured, a chromatogram may not be able to be assembled from the full scans, as there are fewer datapoints. These principals apply to both metabolomics analyses as well as lipidomics. For metabolomics, ionization is almost exclusively done with electrospray ionization, whereas lipidomics sometimes employs atmospheric pressure chemical ionization or APPI that accommodate the greater hydrophobicity of lipids.[29, 81-82] APPI and APCI both pneumatically aerosolize the LC effluent inside the ionization source. APPI ionizes the analytes by using an UV source which has a strong ionization efficiency for hydrocarbons, and APCI uses a corona discharge to ionize analytes which

generally form similar ions to ESI, while having a greater chance of adducts and in-source fragmentation.

Shotgun Lipidomics Details

In lipidomics, mass spectral analysis can be performed without being coupled to chromatographic separation. This "Shotgun Lipidomics" was developed primarily through the work of Han and Gross.[27-28, 83] In shotgun lipidomics, lipid extracts are analyzed with mass spectrometry by direct infusion without fractionation. The use of collisional induced dissociation (CID) with a triple quadrupole instrument enabled Han and Gross to identify both the lipid headgroup and corresponding tails with greater than 95% accuracy and quantify with r-squared values of at least 0.99.[27]

Perhaps surprisingly, shotgun lipidomics can also be used non-polar/neutral lipids. Han and Gross have also pioneered methods for the direct infusion ESI-MS/MS analysis of triacylglycerols (TAGs). This may be surprising because nonpolar lipids often have low ionization efficiencies, and solvents often used for neutral lipids are aprotic and do not spray well. To account for this, Han and Gross used a 50/50 chloroform:methanol solution that fully solubilizes the TAGs while still spraying in ESI.[83] To facilitate ionization, lithium hydroxide was added forming lithiated TAG that ionizes in positive mode.[83-84] This process allows not only for qualitative fingerprinting of TAG species but also quantitative analysis.[83]

Tandem mass spectrometry with full-scan high resolution instruments such as Q-TOF and QE allow for head and tail profiling as well, but faster and with more information (each MS^2 scan captures all product ions).[85-86] The improved power has been demonstrated with a shotgun lipidomics method employing both a tandem orbitrap and a Q-TOF that has been used to quantify 250 different lipids of 21 classes in yeast.[87] This method developed by Ejsing et al. utilizes a two-step biphasic extraction to capture polar and nonpolar lipids. They analyze this extract via nano-ESI direct infusion and quantify lipids from the classes PC, LPC, PE, PG, PA, PS, LPS, PI, LPI, CL, LCBP, IPC, MIPC, $M(IP)_2C$, DAG, TAG, LPC, and LCB. They estimate a 95% lipidome coverage of their test species (*S. cerevisiae*). These are impressive results and display a robust methodology. However, the method uses multiple analysis methods which throttles throughput and low abundance lipids (which may be involved in signaling and the most biologically relevant) will be missed.

Informatics

Data Visualization

All these different methods only process and measure samples. Once the measurements have been made, the instrument will output a matrix of retention time (when chromatography is employed), *m/z*, and intensity. This gives no biological information. How does one link the MS results to metabolites and explain biological relevance? Informatics. Many different informatic and statistical approaches are needed to interpret the vast amounts of data generated by typical metabolomics experiments. The approaches used vary depending on the experiment type. Targeted metabolomics experiments can often be interpreted by traditional analytical chemistry approaches. These include calibration curves, standard deviations, and the like. As for non-targeted metabolomics, the task becomes much more difficult. Modern orbitrap and TOF instruments generate high-resolution full spectrum measurements multiple times a second. If the entirety of these scans taken across a thirty minute (or longer) chromatographic run are to be considered, there is far too much data to be annotated by hand. To begin the analysis, peak picking algorithms such as XCMS are often employed. XCMS (standing for any type of chromatography-mass spectrometry) was developed by the Suizdak lab for automated peak picking and non-linear retention time correction of metabolomic and lipidomic data sets.[88] The benefits of peak picking are straight forward, allowing for the selection of thousands of features in a short period of time, but the benefits of non-linear retention time correction may not be intuitive. Across a sample set, retention times vary and drift differently for each analyte. One compound may have a longer retention time while a second has a shorter retention. This is especially true when the samples vary in composition, which can often be the case with extracted samples. Regardless of the cause, XCMS is able to correct for these unpredictable shifts and output high-quality lists of spectral features with minimal effort. XCMS is now available as an online open access program that allows for rapid annotation of chromatography-mass spectral data.

After generating a list of possible metabolites, the task of parsing through and distinguishing information from noise can begin. The first steps usually taken are to try to identify adducts or isotopes and remove or consolidate them to ensure one peak per compound. This is usually done by a taking a peak and looking for features that are aligned in the time domain but shifted in the mass domain by a known amount (the mass of an adduct or isotope).

There are different scripts that can perform this task including CAMERA,[89], MetAssign,[90] or AStream.[91] This consolidation helps reduce the amount of data to be analyzed and greatly reduces redundancy, which can skew interpretation, particularly for multivariate analyses like principal component analysis (PCA) or partial least squares-discriminant analysis (PLS-DA). These statistical methods are helpful in describing relationships between sample sets. These attempt to discern relationships between groups by reducing the data to eigenvectors, which describe the variance between samples. This is a good way to visualize the differences between several samples over many data points. Without reduction of dimensionality in this way, such differences would be obscure to incomprehensible. Although similar, PCA and PLS-DA differ in significant ways. Primarily, PCA is an unsupervised analysis, whereas PLS-DA is an unsupervised analysis. This means that PCA attempts to plot whatever major differences exist between input samples, whereas PLS-DA plots the differences between samples so as to maximize the covariance between assigned groups. Practically, this means that PCA is not as sensitive as PLS-DA is to subtle differences between groups. PLS-DA, on the other hand, can be prone to overfitting. Overfitting is when the model finds and overemphasizes small differences between groups that may just be due to chance. So, PLS-DAs must be used carefully to avoid this issue. Beyond qualitatively distinguishing between sample groups, these are useful in hypothesis generation as well. The eigenvectors, which show separation between groups, can be deconstructed and "loadings" (weighted contributions to the separation) can be generated for each variable (spectral feature or analyte). Analytes with large loadings show significant deviation between groups and may be important features associated with the test variable. These analytes might form the basis of a future targeted analysis to validate and elaborate what the meaningful differences might be. Heatmaps are also helpful in this regard. These graphically display the foldchange associated with different features and will sometimes also display statistical significance in the form of a T-test p-value. This has the advantage over PCA and PLS-DA of displaying magnitude as well as statistical significance in feature variation.

MAVEN is an opensource software for metabolomics data processing that attempts to integrate the preprocessing with visualization, one of the first to do so.[92] This can be operated in a fully automated process, useful for non-targeted metabolomics, or manually by imputation of masses or standards lists and either manual or automatic integration of extraction ion chromatogram (EIC) peaks. It utilizes a retention time alignment mechanism similar to XCMS.

Visualization is facilitated by the automatic generation of EICs for features. Additionally, isotope peaks can also by generated and included in the data output through a simple GUI.

Unknown Identification

It is rare for spectral features to be easily assignable to specific metabolites. Often features are ambiguous and even when they are annotated by software, these assignments are putative and not always reliable. *De novo* identification of these features is a very complicated task. The first step is generating possible formulae from the exact mass. There are various programs capable of doing this. Normally, some filtering occurs at this step. This is because some formulae that may mathematically match the observed mass are obviously false. For example, both $[C_6H_{12}O_9P]^-$ and $[H_{23}BrNO_5SP]^-$ may fit an observed 259.0218 *m/z* (with less than 1 ppm mass accuracy); however, the latter can be rejected as highly improbable. Therefore, generators will commonly apply constraints based on unsaturation or limitations on the possible number of uncommon elements. However, a more rigorous filtering can be applied to narrow the options further. Such a process was first described by Kind and Fiehn in 2007.[32] They employ six or seven (depending on the type of analysis) metrics to filter mathematically possible formulae, the so called "Seven Golden Rules." Two of these seven rules have already been mentioned, restrictions on element numbers and ratios. A third rule employed attempts to determine the ion adduct detected (typically $[M+H]^+$ in electrospray) and then determine if the neutral compound is stable. This would filter out species based on unbalanced valences (improbable radical species for instance). A fourth rule is an isotope pattern filter. This attempts to narrow the possibilities by eliminating formulae that do not correspond to the observed isotope pattern. This can be particularly determinative for elements with unique isotope distributions such as chlorine, bromine, and sulfur. Rule five applies a heteroatom to carbon ratio check as this will be a small number for most compounds. Rule number six is used as a follow-up to element ratio check where formulae with several high element ratios are filtered. The example given is formulae like $C_{26}H_{28}N_{17}O_1P_3S_8$, which "would pass all rules so far including the element ratio check; however, the combination of high element ratios would still be too improbable."[32] Finally, the seventh rule is a trimethylsilyl (TMS) group check. TMS is a derivatizing agent used to make polar compounds volatile. This is used often with GC-MS metabolomics, but this rule is not used outside of that niche. These rules eliminate 92% of the hypothetically possible formulae (with

only C, H, N, S, O, and P) below 2000 Da. Additionally, for high quality spectra, the compound can be correctly mapped to compounds existing in a database with 98% accuracy (larger databases like PubChem show lower accuracy but still >88%), see Figure 1 for a graphical representation. It is worth noting that the results in Figure 1 are based on cumulative FPR, and so the actual accuracy for high mass compound is less than displayed by the graph. Regardless, the seven golden rules paradigm is very useful and lays the groundwork for further improvements to shrink the possible formulae even more. Many current formula generators incorporate at least some of the rules described here.

A more common method of metabolite identification is simply to compare *m/z* values (combined sometimes with isotope intensities, retention times, or MS^2 spectra) to libraries. This is facilitated by high mass accuracy instruments as the better the accuracy the better the library matches. There are various databases, including the FiehnLib already mentioned, but also MetaCyc, Lipid Maps, MassBank, KEGG, Human Metabolome Database, and even PubChem can be searched.[16, 33, 93-98] This is very powerful as a high-throughput method of metabolite identification; however, since most of these are not easily searched with MS^2 data and retentions times are rarely meaningful unless the method is copied exactly, such searches are often restricted to a sheer exact mass matching. This performs better than merely calculating possible formulae, but it still is powerless to resolve isobaric metabolites which can be common.

Biological Interpretation

These processing steps of generating and validating an annotated list of features are really pre-steps to facilitate biological interpretation. Although this is the ultimate goal of the analysis, it can take different forms. Due to the rapidly shifting nature of metabolism, the most important factor is reproducibility/statistical significance. Vast metabolite pool size changes can occur as a result of factors as simple as different delays before quenching. Care must be taken to filter irreproducible features as these are likely artifacts of the labile nature of metabolism. When

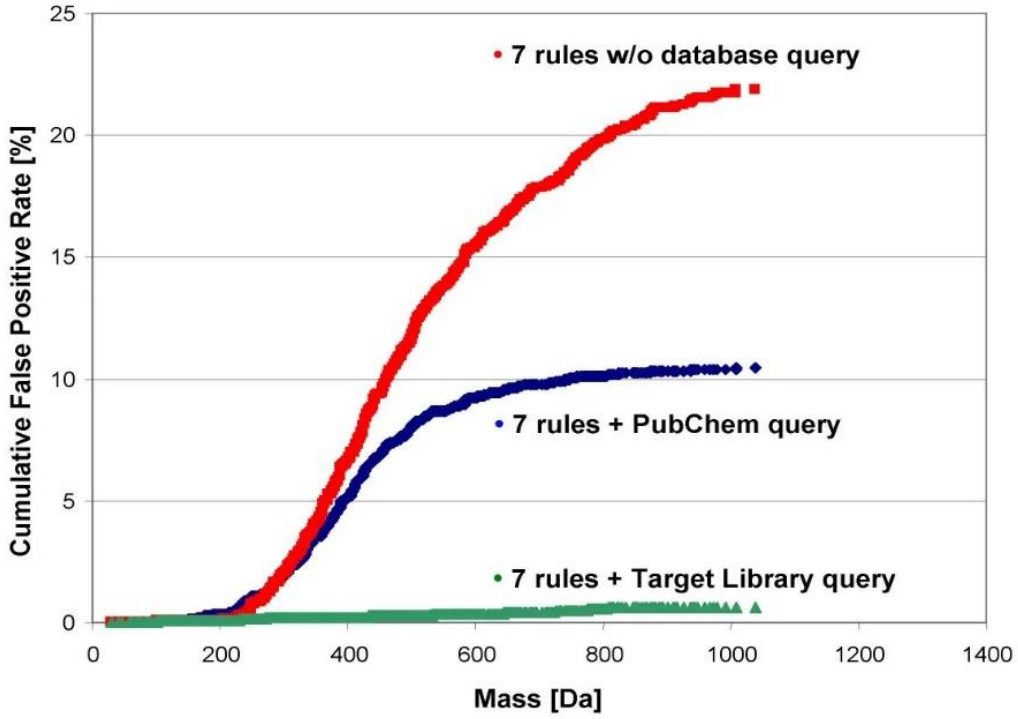

Figure 1-1. Seven Golden Rules False Positive Rate. False positive rates for seven golden rules analysis of 2,400 DrugBank molecules with a simulated ±3 ppm accuracy and ±5% isotope ratio error based on compound mass. The red curve shows the cumulative FPR (when the correct assignment is not in the top 3 results) for formula assignment without database query. The blue and green curves show cumulative FPR (considering only the top result) for formula assignment with database query, PubChem and DrugBank, respectively. Reproduced from ref[32] under the CC BY 4.0 license.

biomarker discovery is desired, the biological analysis can be as simple as generating a volcano plot or its equivalent. A volcano plot plots the fold-change and statistical significance (p-value) of each feature relating a test group to a control. Potential biomarkers are features with both high fold changes and significance and are easily identified by such a graph.

For a more detailed biological analysis, there are different options such as MetaboAnalyst.[99] MetaboAnalyst is a free program that can take metabolite inputs and describe the pathways involved in the observed changes. This can be illuminating depending on the experiment; for example, a pathway suppressed in a gene knockout could indicate that gene encodes an enzyme in the pathway or else is involved in its regulation. MetaboAnalyst allows for the integration of other data sets including genomics or multiple metabolomics experiments. This program is convenient and user-friendly interface and facilitates analysis.

Unknown metabolites are more challenging. If the metabolites demonstrating the greatest changes from a treatment are unannotated, it is not possible to say what metabolomic networks they are affecting. The simplest way to treat these is to designate them as biomarkers and analyze them as discussed previously. Biomarkers are biological signals that correlate with a biological state and can be used as a diagnostic tool. This has been done in several metabol/lipidomic studies for diseases including traumatic brain injury, Alzheimer's disease, and cancer.[100-102] However, outside of disease profiling experiments, biomarkers are not relevant and other techniques must be employed to extract meaningful biological conclusions. There has been some interest in using van Krevin diagrams to visualize non-targeted metabolomics data.[103] Van Krevelen diagrams plot features with assigned formulae based on their carbon to hydrogen and carbon to oxygen ratios. This can be helpful because different classes of compounds tend to have certain ratios. This allows features to be sorted by their probable metabolite class, allowing some speculation into biological function. Figure 2 displays metabolomics data visualization using a van Krevelen diagram.

Specific Contributions

In this diverse and complicated field, my research has focused on two areas. The first, is the application of metabolomics in a multi-omics experiment to expand the scientific knowledge base of metabolism. Specifically, chapter two describes an integrated metabolomics and genomics experiment done in collaboration with Dr. Kania in the Department of Biomedical and

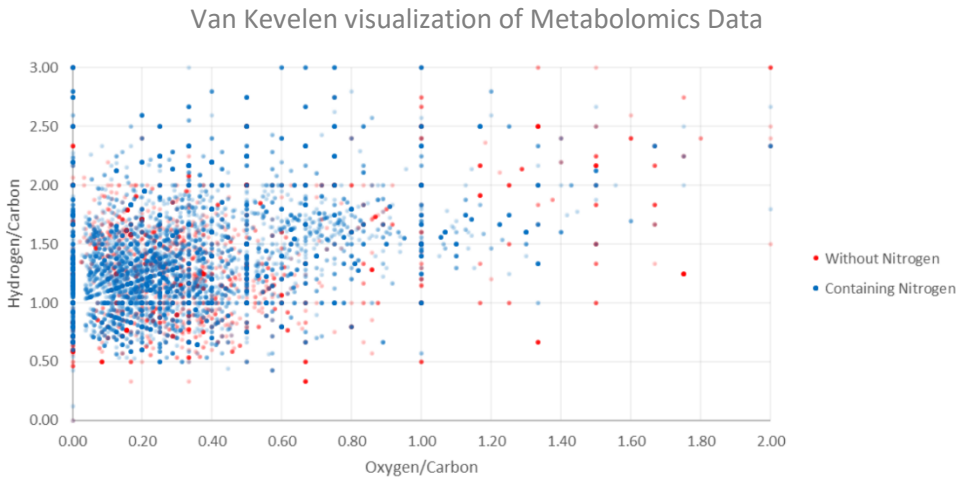

Figure 1-2. Utility of van Krevelen Diagrams for Metabolomics Visualization. A visualization of metabolomics data using a van Krevelen diagram. This was generated using the program OpenVanKrevelen.[103] Metabolites are plot based on their expected hydrogen to carbon and oxygen to carbon ratios. Blue dots also have nitrogen in the predicted formula, while red dots do not. The opaque dots have more features associated with them.

Diagnostic Sciences to characterize the metabolic pathways of clinically relevant *Staphylococcus* isolates. This is an example of a metabolomics experiment which helps to shrink the knowledge gap by applying established metabolomic tools to elucidate the metabolic web while guiding medical treatment of potentially lethal bacterial infections.

The second area is in the advancement of metabolomic techniques for better multiomic analyses of bacteria. Chapter three describes work done to validate different multiomic extraction techniques and critically evaluate them based on their extraction efficiency. These were tested on their ability to collect metabolites, lipids, and proteins in a simultaneous extraction, and found to be viable for multi-omics applications. This was a collaborative work with Dr. Hettich in the Microbiology Department and is important to guide future multiomic experimentation by providing a strong extraction protocol that will conserve sample, save time, and give more accurate results. These both demonstrate metabolomics is a varied and multifaceted field and display the power and complexity of multiomic integrated metabolomics.

Chapter 2: Metabolomics Applied: A metabolomic and genomic investigation of *Staphylococcus* isoprenoid synthesis

This chapter contains material that is in preparation to be included in a future publication titled: Whole Genome Sequence and Comparative Genomics and Metabolomics Analysis of Human and Canine *Staphylococcus schleiferi* isolates. The proposed authors are Mohamed A. Abouelkhair, Matthew J. Keller, Hector F. Castro, Shawn R. Campagna, and Stephen A. Kania. SAK, MAA, SRC, and HFC contributed to the conceptualization of the project and experiments. Experiments were conducted by MAA and MJK. Formal analysis was performed by MAA, MJK, and SAK. Writing of the original manuscript was done by MAA and MJK with edits by MAA, MJK, SRC, and HFC. This has been adapted more fully for this thesis.

Introduction

Bacterial species of *Staphylococcus* are one of the most common causes of disease. In 2017 alone there were 119,000 bloodstream *Staphylococcus* infections (not including skin infections) which led to 20,000 deaths.[104] Many of these infections are thought to be caused by *Staphylococcus aureus*. However, there is growing evidence that other species of *Staphylococcus* are often misidentified as *S. aureus*. Two species in particular, *S. schleiferi* and *S. pseudintermedius*, are more commonly associated with disease in canines but been shown to also cause disease in humans.[105-114] Both *S. schleiferi* and *S. pseudintermedius* are easily (and potentially often) mistaken for *S. aureus*.[114-115] Antibiotics are an important defense against *Staphylococcus* infections, however, antibiotic resistance presents problems with methicillin-resistant strains of *S. schleiferi* and *S. pseudintermedius* as well as the more common methicillin-resistant *S. aureus*.[113, 116-117] These species, then, are a danger to public health and challenge health care as better diagnostic and treatment options are needed.

Isoprenoid biosynthesis is one of the essential metabolic pathways and is very similar between bacteria, eukaryotes, and plants.[118-119] Different species of the staphylococcus genus are known to use different pathways of isoprenoid biosynthesis. *S. aureus* has been shown to use the mevalonate pathway for isoprenoid synthesis while others such as *S. schleiferi* and *S. pseudintermedius* have been shown to use the non-mevalonate pathway.[120-121] This metabolic difference was proposed to be related to the host species as a critical intermediate in the non-mevalonate pathway triggers an immune response in human and primate hosts and the

mevalonate pathway could be a virulence factor in *S. aureus*.[121] Additionally, fosmidomycin, a phosphonic acid derivative that has been used to target the non-mevalonate pathway via inhibition of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (Dxr) is identified as a promising antimicrobial to specifically treat staphylococcus species infecting animals.[121]

This problem is one that is naturally suited for metabolomics investigation, the mevalonate and non-mevalonate pathways have several unshared intermediates which should give a relatively easy way to distinguish the metabolic pathway used. In this experiment a semi-targeted metabolomic analysis is employed to characterize the pathway used and comparative genomics is used to go deeper into the differences between the bacterial species and confirm metabolomics findings. This demonstrates the synergistic effect of using multiple omics techniques for biochemical investigations.

Materials and Methods

Bacterial Strains, Media and Growth Conditions

Bacteria propagated in this study included five *S. schleiferi* isolated from human cases in the USA (191, 192, 196, CDC:132-96, CDC: 78-04) and the *S. schleiferi* subsp *schleiferi* type strain ATCC43808 isolated from a human patient in France. A single bacterial colony of each strain grown on blood agar plates was inoculated into 5 mL of sterile tryptic soy broth (TSB) (BD Biosciences, USA; Cat. no. RS1-011-21) and incubated overnight at 37 °C with shaking at 225 rpm.

DNA Extraction, Library Preparation and Whole Genome Sequencing

DNA extraction was performed using the MasterPure DNA purification kit (Epicentre, USA; cat. no. MCD85201) according to the manufacturer's instructions. Libraries for sequencing were prepared using the Nextera DNA sample prep kit (Illumina, Inc., USA) according to the manufacturer's instructions. The genomes were sequenced using a MiSeq platform (Illumina, Inc.) with two runs (75 bps forward and reverse) at the University of Tennessee Genomics Core facility. Sequences were trimmed using BBDuk and de novo assembled using Geneious Prime® 2019.0.4.[122] A quality assessment tool for genome assemblies (QUAST) was used to assess the quality metrics of the assembled genomes [123]. Annotation was performed by the NCBI Prokaryotic Genome Annotation Pipeline version 4.6

(https://www.ncbi.nlm.nih.gov/genome/annotation_prok) using the best-placed reference protein set with GenMarkS+.

Comparative Genomics Analysis

For comparative genome analyses, five canine *S. schleiferi* isolates (*S. schleiferi* 1360-13, *S. schleiferi* 2142-05, *S. schleiferi* 5909-02 and *S. schleiferi* 2317-03 with accession numbers of CP009470, CP009762, CP009676 and CP010309, respectively) were used in addition to *S. pseudintermedius* HKU10-03 (NC_014925.1), *S. pseudintermedius* ED99 (NC_017568), *S. lugdunensis* HKU09-01 (CP001837), *S. lugdunensis* N920143 (FR870271.1), *S. epidermidis* ATCC 12228 (NC_004461), *S. epidermidis* RP62A (NC_002976.3), *S. aureus* subspecies *aureus* ST398 (NC_017333), *S. aureus* subspecies *aureus* USA300_FPR3757 (NC_007793), *S. aureus* subspecies *aureus* COL (NC_002951), and *S. aureus* subspecies *aureus* NCTC 8325 (NC_007795). A circular graphical display of the distribution of the annotations in each human *S. schleiferi* genome was performed using Circos [124], then whole genomes were aligned using the Basic Local Alignment Search Tool (BLAST, v. 2/2/22) .[125] Pan/core-genome sizes were computed using the MicroScope gene families (MICFAM) based tool which uses an algorithm implemented in the SiLiX software (http://lbbe.univ-lyon1.fr/-SiLiX-.html): a single linkage clustering algorithm of homologous genes sharing an amino-acid alignment coverage and identity above 80%. Core-CoDing sequence (CDS), variable-CDS and strain specific sizes were determined.[126] Phage prediction was performed using PHAST (**PHA**ge **S**earch **T**ool) (available at http://phast.wishartlab.com/).[127] The presence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) was evaluated as had been done previously.[128]

Functional gene categories were determined with the Rapid Annotation using Subsystem Technology (RAST) v. 2.0.[129] Metabolic pathway reconstructions of each strain were compared using the terpenoid backbone biosynthesis pathway from KEGG.

The nucleic acid sequence of *sodA* from 20 staphylococcal species were aligned using the clustalW algorithm implemented in the software Geneious Prime® 2019.0.4 [122] and the phylogenetic tree was constructed using the tree-building algorithm Neighbor-Joining with the Jukes-Cantor distance estimator implemented in Geneious Prime® 2019.0.4. The *sodA* gene sequence from *Macrococcus caseolyticus* was set as the outgroup.

Metabolic Pathway Analysis

To elucidate the metabolic pathway used for isoprenoid synthesis in different strains of *Staphylococcus*, relevant metabolites were analyzed from cellular extracts to compare pool size. Two isolates of *S. aureus* were tested as well as five *S. schleiferi* isolates and one *S. pseudintermedius*. Isoprenoid metabolites were analyzed, including isopentylpyrophosphate (IPP), 2-*C*-methyl-D-erythritol-phosphate (MEP), mevalonate, phosphomevalonate, and geranyldiphosphate (GPP). These were examined using an established metabolomic method. This method was performed in a "semi-targeted" manner in that metabolites were annotated based on exact mass and retention time compared to a standard library. This was not a full non-targeted method as features without standards were not investigated (excepting phosphomevalonate), but was not a true targeted experiment as the instrumental system was not biased toward detection of particular compounds, nor was analysis limited to a subset of metabolites with internal standards.

Standards were purchased for three metabolites involved in isoprenoid biosynthesis. These were isopentylpyrophosphate (IPP), 2-*C*-methyl-D-erythritol-phosphate (MEP), and mevalonate. Mevalonate (pn: 42147) was purchased from Sigma as an analytical standard. MEP (pn: 52131) was purchased from Sigma but was reagent grade. IPP was purchased from Fisher (pn: I00501MG) also as reagent grade. Reagent grade was considered to be acceptable since the standards were used only to establish retention times and not for quantification. Geranyldiphosphate had been previously annotated using this method and no standard was used in this experiment.

Extraction of Metabolites

For the metabolomics experiment isolates *S. schleiferi* 192, *S. schleiferi* 196, *S. schleiferi* 132-96, *S. schleiferi* 182159, *S. schleiferi* 182116, *S. pseudintermidus* 06-3228, *S. aureus* USA300, and *S. aureus* ST398 were used. Metabolites were extracted from bacterial cultures using a modified method based on a procedure by Rabinowitz and Kimball.[53] 5 mL of each culture were vacuum filtered through nucleopore polycarbonate filters to collect cells (Whatman, Little Chalfont, U.K.). The cultures were analyzed in biological triplicate. The filters were then placed cell side down into petri dishes containing 1.3 mL of extraction solvent (40:40:20 HPLC

grade methanol, acetonitrile, water with 0.1% formic acid). The solvent and dishes had been pre-chilled in a –20 °C freezer while the filtration was set up. The filters in the solvent were placed at –20 °C, to facilitate extraction of metabolites, for 20 min. The following steps were completed in a 2 °C cold room. Filters were rinsed with the extraction solvent, and the suspension was transferred into a 2 mL centrifuge tube. An additional aliquot of extraction solvent (400 µL) was used to wash the filters, and this was added to the other 1.3 mL aliquot in the centrifuge tubes. The tubes were centrifuged at 13,000 x*g* for 5 min, and the supernatant was transferred to a second tube. A further aliquot of extraction solvent (200 µL) was used to resuspend the remaining cell pellet, and this suspension was allowed to extract at –20 °C for 20 min. The resulting supernatant was collected via centrifugation as explained above, and this second extraction was added to the first. All samples were then dried under a stream of nitrogen (there was some sample loss at this step due to the nitrogen spray splashing samples). The dried material was stored at –80 °C before thawing and resuspension in HPLC grade water (300 µL) for ultra-performance liquid chromatography—high resolution mass spectrometry (UPLC-HRMS) analysis.

Instrument Parameters

Samples were analyzed with an established metabolomics method[35, 130] using an Ultimate 3000 UPLC (Dionex, Sunnyvale, CA) coupled to an Exactive Plus Orbitrap (Thermo Scientific, Waltham, MA). Briefly, 10 µL of each sample were injected onto the UPLC and separated with a Synergi Hydro-RP column (Phenomenex, Torrance, CA) using a gradient of 97:3 water/methanol containing 15 mM acetic acid and 11 mM tributylamine as an ion pairing reagent (solvent A) with pure methanol (solvent B). The gradient was as follows: from 0 to 5 min solvent B increased from 0% to 20%, from 5 to 13 min B increased from 20% to 55%, from 13 to 15.5 min B increased from 55% to 95% and was held constant. B then decreased to 0% at 19 min and was held until 25 min. The MS was operated using electrospray ionization in negative ion mode (3 kV spray voltage) with a resolution of 140,000. The automatic gain control was set at 3×10^6 with a maximum injection time of 100 ms and the s-lens RF level was set to 50. The sheath gas flow was 25, auxiliary gas 8, and sweep gas 3 (all arbitrary units). The scan range was 72 to 800 *m/z* for the first 9 minutes and then 110 to 1000 for the remaining 16 min.

Data Analysis

MS convert[131] was used to convert data from RAW to mzML file types[132]. Spectral features were evaluated using the open source software MAVEN[92] and were assigned to metabolites based on retention time and exact mass. Metabolites where none of the sample intensities were 3x greater than the media blank were eliminated from further analysis. Metabolite intensities were then normalized using the optical density at 600 nm for each culture. [130], and p values calculated using the student's T test.

Results

Metabolite Standards

When the standard of isopentylpyrophosphate (IPP) was analyzed using the UPLC-HRMS method, only the monophosphate form was detected on the mass spectrometer, not the pyrophosphate. Given the instability of pyrophosphates, is not surprising that IPP would hydrolyze on column (or in stock solution) into isopentylmonophosphate (IMP). This IMP peak was assigned to IPP and analyzed as such.

GPP was below the detection limit and was not measured. IPP was detected (as the IMP ion) with low intensities in the samples near or at background and so was dropped from further analysis. MEP was measured, but had noisy, low intensity peaks which limit the conclusions that can be drawn. Mevalonate had intense peaks matching the standard's retention time. Phosphomevalonate was putatively identified without a standard. A distinct peak at *m/z* 227.0327 was observed in the samples (corresponding to the phosphomevalonate $[M–H]^-$ exact mass with a 2.8 ppm accuracy). The C13 isotope peak of the feature was about 6.7% of the parent peak. This is consistent with phosphomevalonate having six carbons in its formula. Searching phosphomevalonate's neutral mass on the Human Metabolome Database (HMDB) with a 20 ppm window gives only 5-phosphomevalonate as a possible metabolite.[97] The mass spectral data, then, is consistent with the 227.0327 *m/z* peak assignment as phosphomevalonate and there seem to be few other biotic options that fit the mass.

Metabolomic Results

The metabolomics run generated data for 91 metabolites that correspond to a previously run standard's exact mass and retention time. The full list can be seen in Table A-1, A-2, and A-

3 (see appendix). For the isoprenoid pathway metabolites, the results can be seen in Figure 2-1. There were no large differences between the different strains in the 2-*C*-methyl-D-erythritol-phosphate (MEP) pool size observed, however strains 192 and 214 were somewhat lower than the other strains. The variations were not overly high (all relative standard deviations (RSDs) below 30% except for USA 300 with 36%).

Mevalonate and Phosphomevalonate were both orders of magnitude higher in the *S. aureus* isolates than in the *S. schleiferi* or *S. pseudintermedius* isolates. These differences between groups were statistically significant (p values <0.003 for mevalonate and p-mevalonate). Mevalonate also showed low variation between biological replicates (RSD ≤ 23%), except for strains 192 and 132-96 where the RSD was 55% and 57% respectively. RSDs of mevalonate and phosphomevalonate are especially low for both *S. aurues* strains. Given the high intensities, detector saturation could play a factor in this, however the C13 isotope peaks followed a very similar pattern as did the monoisotopic peaks, indicating that this pattern reflects the concentrations present in the isolates.

Genomic Features of Human *Staphylococcus schleiferi*

The genome size, GC content, predicted coding sequences and predicted RNAs of the five *S. schleiferi* isolated from human cases in USA (191, 192, 196, CDC:132-96, CDC: 78-04) and the *S. schleiferi* subsp *schleiferi* type strain ATCC 43808 isolated from a human patient in France are listed in Table 2-1. A circular graphical display of each genome of human S. schleiferi isolates was constructed to show the distributions of the contigs, CDS, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to known virulence factors (Figure 2-2). A blast atlas where *S. schleiferi* subsp. *schleiferi* ATCC 43808 was used as a reference against which the similarity of 10 other *S. schleiferi* genomes is shown (Figure 2-3). Regions are displayed where there is similarity between the reference genome and one of the related genomes. The plot shows the variation between human *S. schleiferi* (192, CDC: 132-96, CDC: 78-04, 191, 196 and *S. schleiferi* subsp. *schleiferi* ATCC 43808) and canine *S. schleiferi* isolates.

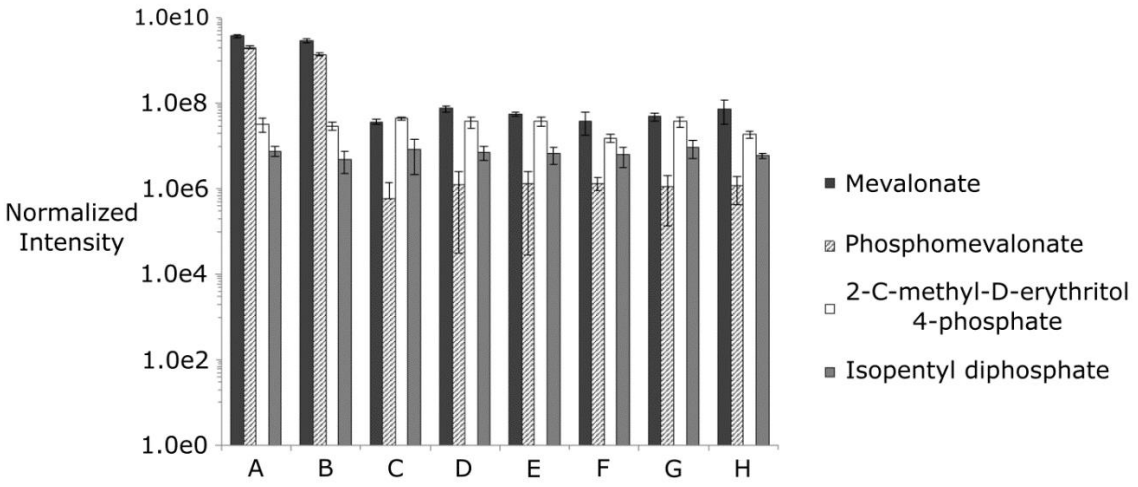

Figure 2-1. Isoprenoid Metabolite Bar Graph. The normalized intensity for each of the detected isoprenoid metabolites are shown for each strain. The biological replicates are averaged with error bars calculated using the standard deviation. The intensities have been set on a log scale to better display the differences. A is *S. aureus* USA300, B is *S. aureus* ST398, C is *S. pseudintermidus* 06-3228, D is *S. schleiferi* 182116, *S. schleiferi* 182150, *S. schleriferi* 192, *S. schleiferi* 196, *S. schleiferi* 132-96. Note that both *S. arueus* strains have high values of mevalonate and 5-phosphomevalonate. The other stains show much lower levels.

Table 2-1: Genomic Features of Human *Staphylococcus schleiferi*. The details of the genomic data for each sequenced S. schleiferi isolate is give. *WGS: whole genome sequence.

Strain	WGS accession No*	No of contigs	N50 (bp)	Genome length (bp)	G+C content (%)	Predicted coding sequences	Predicted RNAs
191	PNRJ00000000	51	138,893	2,508,133	35.73	2,294	72
192	POVG00000000	102	59,786	2,452,487	35.87	2,203	74
196	POVH00000000	56	110,279	2,508,604	35.74	2,299	74
CDC: 132-96	POVI00000000	92	56,958	2,468,342	35.92	2,218	76
CDC:78-04	POVJ00000000	94	57,247	2,469,699	35.92	2,216	76
ATCC 43808^{T}	POVK00000000	88	56,938	2,469,638	35.92	2,218	73

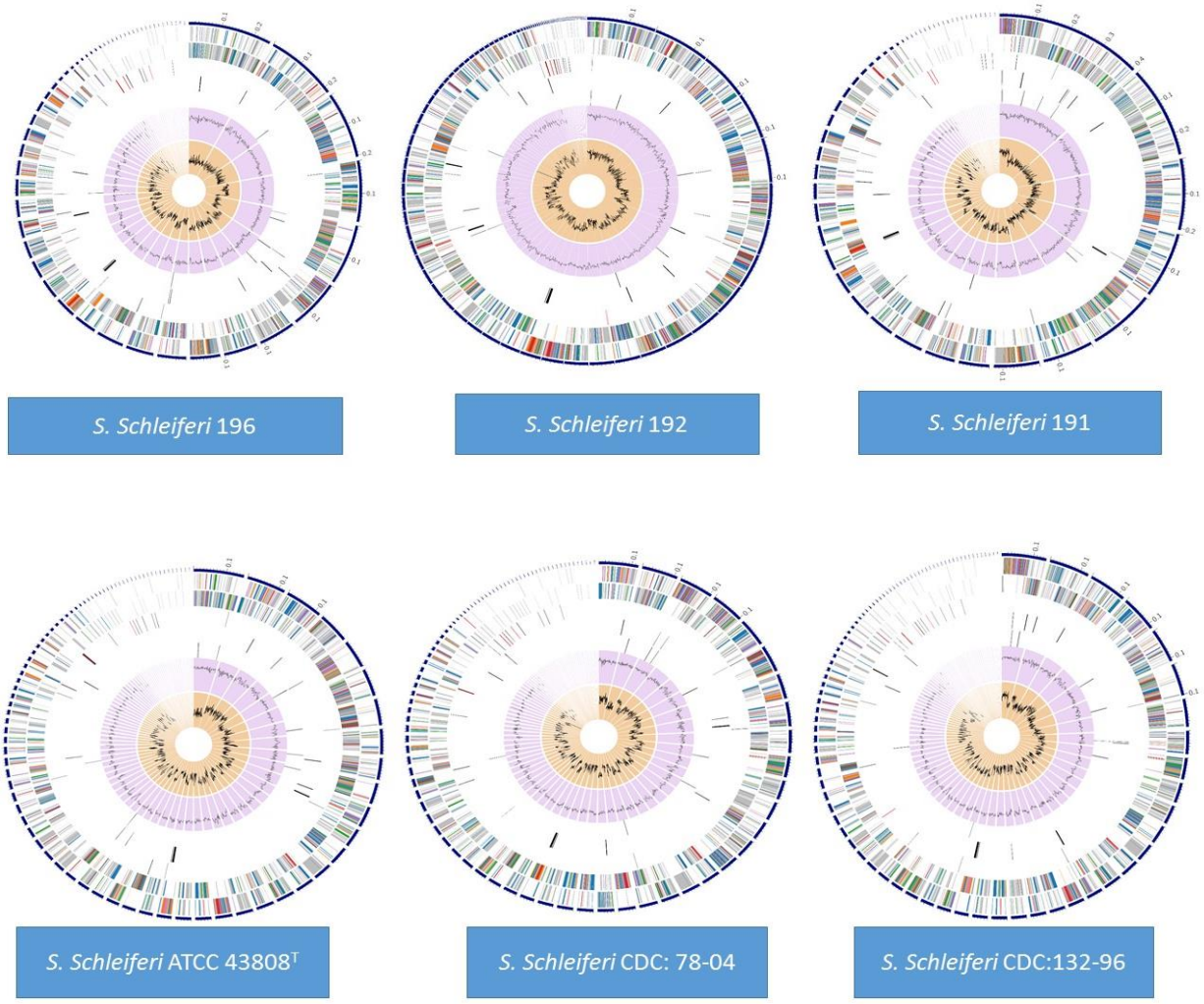

Figure 2-2. A Circular Graphical Display of Human *S. scheiferi* Isolates. This includes, from outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to know virulence factors, GC content and GC skew. The colors of the CDS on the forward and reverse strand indicate the subsystem to which that these genes belong.

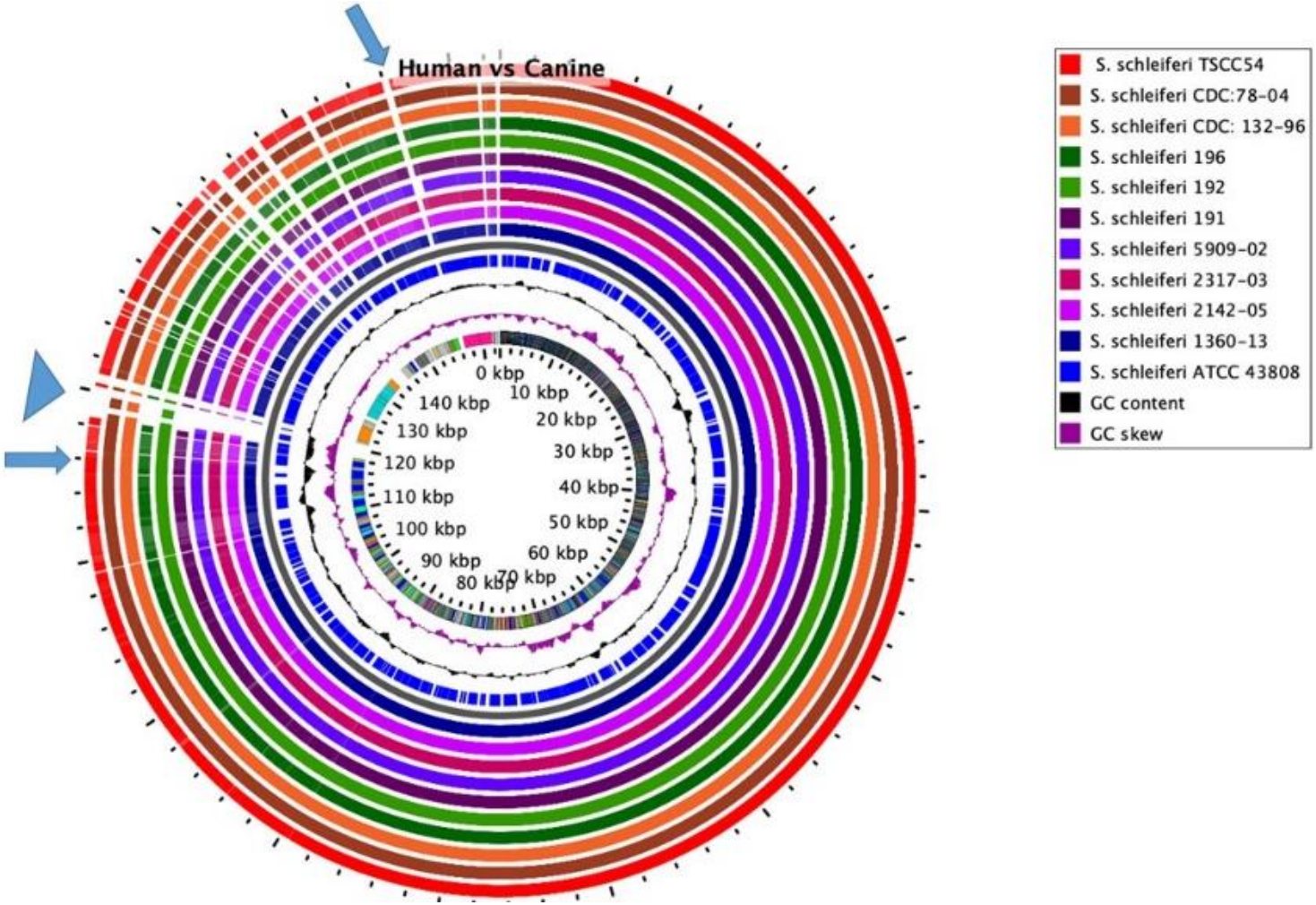

Figure 2-3. *S. schleiferi* Blast Atlas of Human and Canine *S. schleiferi*. The blue triangle and arrows show the variation between human *S. schleiferi* (192, CDC: 132-96, CDC: 78-04, 191, 196 and *S. schleiferi* subsp. *Schleiferi* ATCC 43808) and canine *S. schleiferi* isolates.

Pan/Core Genome of Human *S. schleiferi*

To establish an accurate genomic comparison at the whole genome scale, the pan-genome of the six human *S. schleiferi* isolates was defined using MICFAM with 80% amino- acid (A.A) identity and 80% A.A alignment coverage. The pan-genome includes two distinct constituents, the core and variable genomes. The core genome contains gene families common to all strains while the variable genome is composed of gene families present in at least two strains and absent in at least in one strain. The pan-genome of human *S. schleiferi* isolates consists of 3011 families and 12,268 genes and a core-genome consisting of 2004 families and 10,073 genes whereas the variable-genome consists of 1007 families and 2195 genes (Figure 2-4). The human *S. schleiferi* pan genome was classified into four categories and calculated the number of genes for each genome (Table 2-2) (Pan CDS, core CDS, variable CDS and strain specific CDS).

Further Genomics Data

Phage prediction was performed using PHAST (**PHA**ge **S**earch **T**ool). A complete (intact) prophage (PHAGE_Staphy_EW_NC_007056) was found in *S. schleiferi* CDC: 78-04, *S. schleiferi* 191, *S. schleiferi* subsp. *schleiferi* ATCC 43808 and *S. schleiferi* CDC: 132-96 (Figure 2-5) with sizes of 41.2 Kb, 43.2 Kb, 42.2 Kb and 42.2 Kb, respectively. This prophage contained a minimum of seven hypothetical proteins in all three human *S. schleiferi* isolates. In addition, *S. schleiferi* 191 has an intact 9.4 Kb prophage (PHAGE_Staphy_phiPV83_NC_002486) whereas *S. schleiferi* 196 has incomplete prophages with different sizes.

Presence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) was evaluated for the six human *S. schleiferi* isolates. One CRISPR array with five repeats (average repeat length was 36 nt and average spacer length was 30 nt) was found in *S. schleiferi* subsp. *schleiferi* ATCC 43808, *S. schleiferi* CDC: 132-96 and *S. schleiferi* CDC: 78-04. There are two CRISPR arrays in *S. schleiferi* 191 and *S. schleiferi* 196, the first array consists of seven repeats (average repeat length is 36 nt and average spacer length is 29 nt) and the second array consist of only four repeats (average repeat length is 36 nt and average spacer length is 36 nt). No CRISPR array was identified in *S. schleiferi* 192.

Metabolic pathway reconstructions of each strain were compared using the terpenoid backbone biosynthesis pathway from KEGG and verified in the MicroCyc metabolic database. A hierarchical clustering created 9 clusters based on similar subsystems profiles (a, b, c, d, e, f, g,

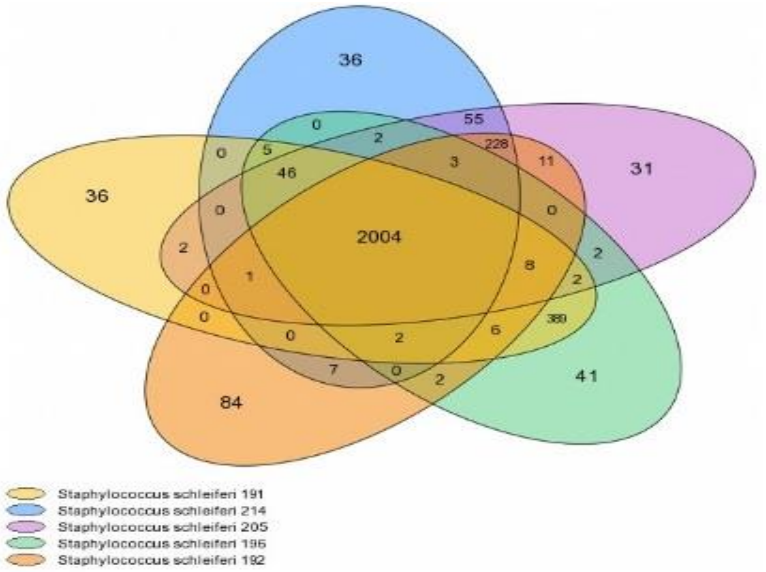

Figure 2-4. Comparison of Shared Genes in *S. schleiferi* Isolates. Venn diagram showing shared and unique genes in human *S. schleiferi* isolates. Numbers inside the circles indicate the genes shared among genomes.

Table 2-2. Pan Genomics CDS Breakdown. List of Pan CDS, core CDS, variable CDS and strain specific CDS count and percent for each human *S. schleiferi* genome.

Organism	CDS	Pan CDS	Core CDS	Var CDS	Strain specific CDS	Core CDS (%)	Var CDS (%)	Strain specific CDS (%)
S.schleiferi 192	2383	2382	2014	368	84	84.551	15.449	3.526
S.schleiferi CDC: 78-04	2414	2411	2017	394	36	83.658	16.342	1.493
S.schleiferi CDC: 132-96	2424	2422	2018	404	31	83.32	16.68	1.28
S.schleiferi 191	2517	2516	2013	503	36	80.008	19.992	1.431
S.schleiferi 196	2530	2529	2014	515	41	79.636	20.364	1.621
S.schleiferi subsp. Schleiferi ATCC 43808T	2421	2419	2014	405	44	83.258	16.742	1.819

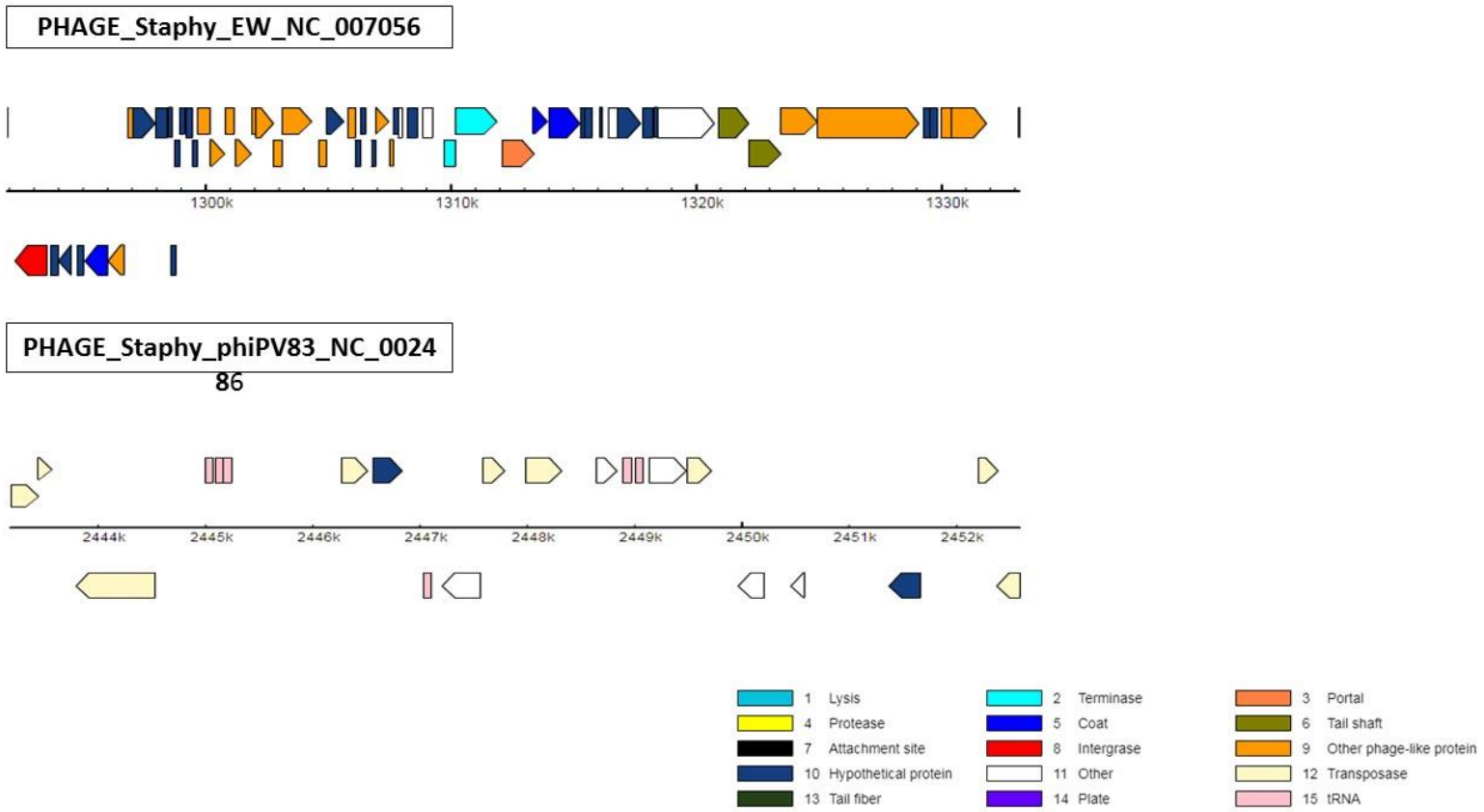

Figure 2-5. Linear View of Human *S. schleiferi* Prophages. Prophage arrangement and position of proteins. Color denotes protein function.

h, j), each cluster showing the subsystems variation that were absent from the group of staphylococcus isolates but present across all other species examined (Figure 2-6**)**.

The Neighbor-Joining phylogenetic tree constructed from the Staphylococcus superoxide dismutase gene (sodA) is shown in Figure 2-7. The *sodA* gene sequence from *Macrococcus caseolyticus* was set as the outgroup. The whole genome sequences of *S. schleiferi* isolated from humans: 191, 192, 196, CDC: 132-96, CDC:78-04 and ATCC 43808^T have been deposited at DDBJ/ENA/GenBank under the accession **PNRJ00000000, POVG00000000, POVH00000000, POVI00000000, POVJ00000000** and **POVK00000000** respectively. The *Staphylococcus schleiferi* 191, 192, 196, CDC: 132-96, CDC: 78-04 and ATCC43808 versions described in this paper are version PNRJ01000000, POVG01000000, POVH01000000, POVI01000000, POVJ01000000 and POVK01000000 respectively.

Discussion

Metabolomic Analysis

Mevalonate and phosphomevalonate were shown to have a much higher pool-size in the *S. aureus* strains than either *S. schleiferi* or *S. pseudintermedius* (the smallest fold change being about 40 for mevalonate and 1022 for phosphomevalonate). Additionally, the mevalonate intensities for *S. schleiferi* and *S. pseudintermedius* are close the value for the media blank ≤ 2.8 × media blank before normalization). Given this significant pool size difference, it seems that the *S. aureus* strains may be using mevalonate metabolically and the other strains are not. Since the main known metabolic role for mevalonate is isoprenoid synthesis, we postulate that isolates *S. aureus* USA300 and ST398 are using the mevalonate pathway for isoprenoid synthesis and strains *S. schleiferi* 196, 192, 132-96, 182150, 182116, and *S. pseudintermedius* 06-3228 are using the non-mevalonate pathway. These are the expected pathways for these species based on previous literature.[133]

Based on the metabolic pathway reconstructions from KEGG and MicroCyc, the canine staphylococcal isolates (*S. pseudintermedius* HKU10-03, *S. schleiferi* 1360-13, *S. schleiferi* 2142-05 and *S. schleiferi* 5909-02) clustered closely together. In contrast, all human *S. schleiferi*, *S. schleiferi* TSCC54 and *S. schleiferi* 2317-03 are segregated suggesting they have a unique metabolic profile as shown in k and g clusters (Figure 2-6**)**. *S. lugdunensis* HKU09-01, *S. aureus* subsp. *aureus* COL and *S. epidermidis* RP62A were found to use the mevalonate pathway for

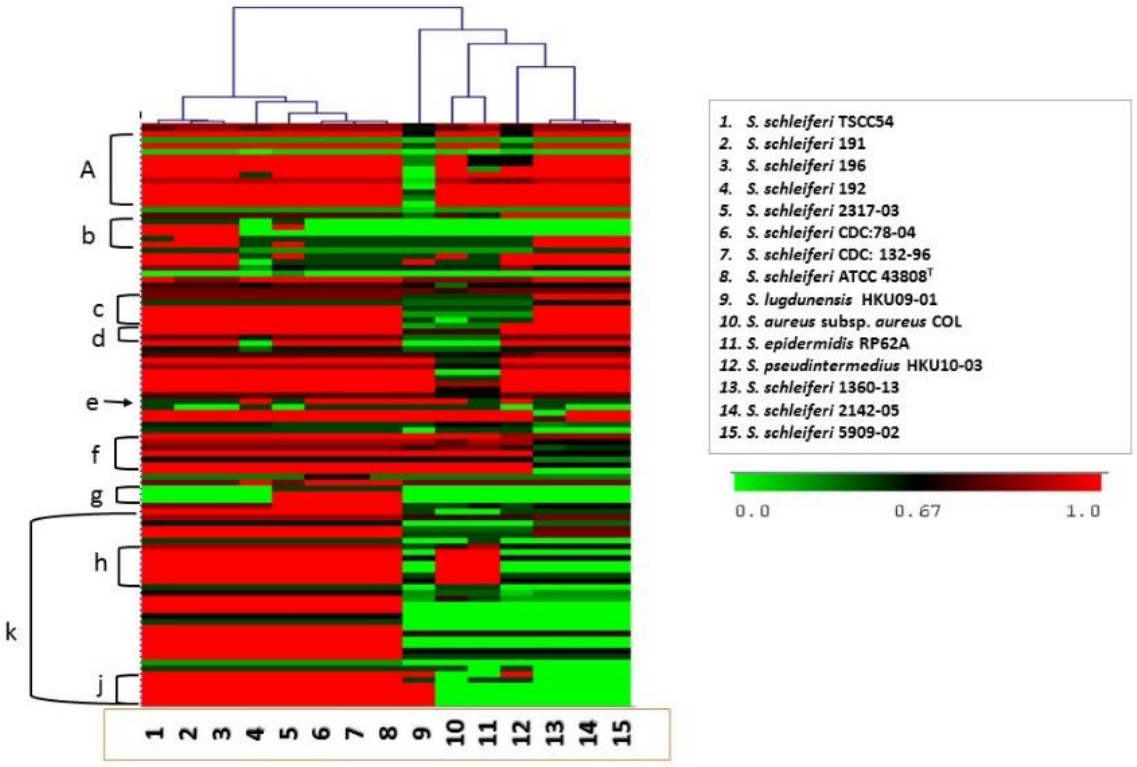

Figure 2-6. Clustering Heat Map of Staphylococcus Isolates. A hierarchical clustering heat map of differentially abundant subsystems among *S. schleiferi* and other staphylococcal species. The top dendrogram shows the relationship between S. *schleiferi* and other staphylococcal species based on subsystems profile similarity.

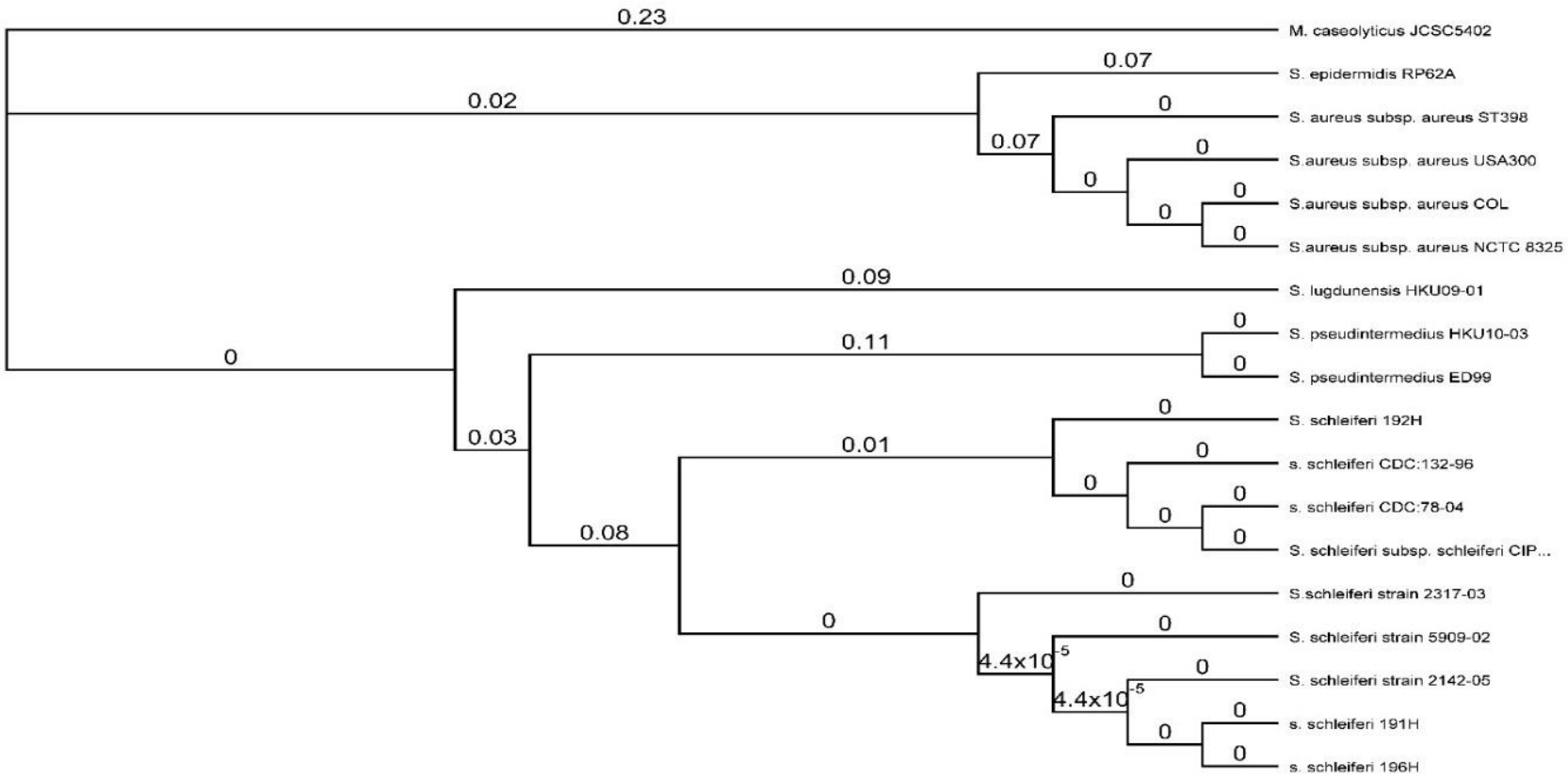

Figure 2-7. Staphylococcus Phylogenetic Tree. A Neighbour-Joining phylogenetic tree constructed from the Staphylococcus superoxide dismutase gene (sodA) is shown. The numbers indicate substitutions per site.

isoprenoid biosynthesis whereas, *S. pseudintermedius* HKU10-03 and all *S. schleiferi* strains examined use the non-mevalonate pathway (2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP/DOXP) pathway) as an alternative method for isoprenoid biosynthesis. This is based on their genomes and is consistent with metabolomics results.

In this study, different species in the Staphylococcus genus were shown to utilize different isoprenoid biosynthesis pathways, which is in agreement with the Misic et al. findings.[134] Furthermore, all of *S. schleiferi* isolates (human and canine) tested in our study were found to use the non-mevalonate pathway as their method for isoprenoid biosynthesis. Mammals use the mevalonate pathway to produce isoprenoids, and therefore the potential of antibiotics to treat bacteria by blocking the mevalonate pathway is precluded. However, for *S. schleiferi* or *S. pseudintermidus* infections, the non-mevalonate pathway is a ready target. Indeed, fosmidomycin has been used in veterinary medicine to treat staph infections in animals by blocking the non-mevalonate pathway.[121] This is not used for human *Staphylococcus* infections due to the prevailing understanding that almost all staph infections are due to *S. aureus*. However, these findings demonstrate that in pathogenic *S. schleiferi* and *S. pseudintermidus* isolated from humans fosmidomycin is in fact a viable treatment option. This should prove especially useful against drug-resistant strains like the already reported methicillin-resistant *Staphylococcus schleiferi*.[121] A hierarchical clustering grouped canine staphylococcal isolates (*S. pseudintermedius* HKU10-03, *S. schleiferi* 1360-13, *S. schleiferi* 2142-05 and *S. schleiferi* 5909-02) more closely together. In contrast, all human *S. schleiferi, S. schleiferi* TSCC54 and *S. schleiferi* 2317-03 have a unique metabolic profile and share some metabolic pathways with staphylococcus species found across human and animal hosts suggesting this is a part of human *S. schleiferi* host specialization. Applying comparative genomics, two complete prophages were identified in human *S. schleiferi*. *Staphylococcus* prophages may play an essential role in the development of bacterial strains and are crucial for the emergence of new virulent *S. schleiferi* lineages.

Conclusions

This experiment utilized semi-targeted metabolomics to elucidate the isoprenoid biosynthetic pathway in *Staphylococcal* isolates. Comparative genomics was also performed on various isolates of *Staphylococcus*, identifying the isoprenoid synthesis genotype, and shedding

more light on the relevance of genetic features such as prophages. Most staphylococcal research has focused on determining the extent of the danger posed by *S. aureus* and the development of new treatments for *S. aureus* infections. However, other species of staphylococci such as *S. schleiferi*, pose a threat to human health and are likely underdiagnosed. The goal of this work was to increase our understanding of *S. schleiferi* metabolism and facilitate future treatment options. The genetic and metabolomic data support each other to accomplish this goal. This information should help guide clinical practice as strains that use the non-mevalonate pathway (which all *S. schleiferi* and *S. pseudintermedius* used) are susceptible to the use of fosmidomycin as an antibiotic. This genomics data should support the research community in other ways as well, such as facilitating identification of pathogenic potential and genetic relatedness of these isolates with *S. schleiferi* isolated from dogs and supporting studies of other bacteria at the genome level. Additionally, it is hoped that this genetic data will support the eventual development of an effective vaccine against staphylococcus infections.

This experiment demonstrates the complementary effect of using multiple omics techniques to build a picture of bacterial physiology and pathology. It is worth noting that the particular combination of metabolomics with genomics was not necessary. Metabolomics combined with proteomics or transcriptomics would also have been successful in characterizing isoprenoid biosynthesis. Whatever the combination, however, by mixing techniques and approaches conclusions can be validated from multiple angles.

Chapter 3: Metabolomics Advanced: An examination of simultaneous extractions for multi-omic analyses

Introduction

While metabolomics itself can be a useful tool, more often these studies are better when integrated with other techniques. One example has already been given in chapter 2. In that study, genomics and metabolomics were used complementarily to investigate the metabolic pathways of *Staphylococcus* isolates. These approaches are becoming more common and various combinations of genomics, transcriptomics, proteomics, metabolomics, and lipidomics show up throughout the literature.[135-136] Often in these omic studies, sample preparation and extraction is one of the most labor-intensive steps. Performing one extraction across multiple samples can be very time consuming, and this is greatly increased if multiple extractions are performed to capture different classes of molecules. Additionally, if extractions are not performed on the same sample at the same time, variations in the physiological state inevitably occur. This is especially true for the metabolomics fraction as the metabolite profile can remarkably change in a matter of seconds.[137] In the case of limited sample amounts (common with clinical experiments), being frugal with sample use is critical. And in some cases (such as tissue biopsies), it is impossible to get more than one sample from the same individual with the same composition (or in the case of continuous studying without killing the individual). For these reasons, if multiomics (such as metabolomics, lipidomics, and proteomics) is employed, a method for the simultaneous isolation of each fraction is extremely desirable. In this chapter, two multiomics extraction methodologies and two lysis procedures are compared and evaluated for extraction efficacy of lipids, proteins, and metabolites.

Chloroform, methanol, and water extractions are commonly used for both lipidomics and metabolomics.[47, 51, 55-58] This is a biphasic extraction that can be conveniently adapted for the collection of both metabolites and lipids. Similar extractions are also used in proteomics preparations to purify proteins and wash out contaminates.[138] This has been used occasionally for the preparation of multiple omics analyses as a concerted extraction.[139-141] One paper in particular used this for the co-isolation of metabolites, lipids, and proteins, the so-called MPLEx method (metabolite, protein, and lipid extraction).[61] Having been extensively tested for metabolites and lipids in very similar methods previously, the authors focused their validation on

the protein analysis and found that although MPLEx resulted in less than one third of the total protein content compared to a control method, they detected approximately the same number of peptides and proteins in their proteomic analysis, and concluded the method was viable for multiomic experiments.

There is another method that has been used for simultaneous extractions. This is a method similar in principal to CME but employing methyl-*tert*-butyl ether (MTBE) instead of chloroform. This has the immediate advantage of avoiding a probable human carcinogen in the extraction solvent. This originated in a paper by Matyash et al. regarding a lipid extraction procedure that was more convenient than CME since MTBE is less dense than water and the organic phase forms the top layer.[62] This also precipitates protein at the bottom of the vial, simplifying fraction collection. Matyash concluded that this method was at least as effective as Folch's CME lipid extraction. This was adapted by Coman et al. as a method entitled: Simultaneous Metabolite, Protein, Lipid Extraction (SIMPLEX).[60] SIMPLEX was found to perform similarly to a control experiment for phosphoproteomics and to give good results for the lipid analysis. The metabolite analysis showed moderately lower intensities as compared to the control and slightly higher relative standard deviations (RSDs). Despite slight drawbacks, the authors concluded that SIMPLEX is a viable option. Although both these methods, MPLEx and SIMPLEX, were tested against dedicated single extractions (the control groups), there has not been a direct comparison of different multiomic extraction methods to date.

Both SIMPLEX and MPLEx employ sonication as a cell lysis step. Lysis is a very important part of any extraction to ensure complete extraction and metabolic quenching. Sonication is an aggressive lysis method that does ensure cell lysis, but results in sample heating and may cause metabolite degradation, especially since the sonication is typically not done with a high organic solvent composition and, therefore, may retain enzyme activity. Other lysis methods have been used as well, including acid, freeze-thaw cycles, high organic solvents, or some combination of these. These different procedures have generated conflicting results and clarifying what methods are effective for maximum extraction with minimal degradation will help guide future experimentation.

Pseudomonas putida was used as the test organism in this project. Strain KT2440 was used, which is a plasmid-free version of a toluene-dregading *P. putida* strain isolated from a field in Japan and is the best characterized saprophytic Pseudomonad.[142-143] This is an exciting

bacterium because of its remarkable potential for bioremediation. Indeed, studies have demonstrated the viability of *P. putida* for bioremediation of diverse pollutants such as petroleum, organophosphates, inorganic cyanides, organic solvents, and halocarbons.[144-147] This also could serve as a model organism, as a safer version of the related *Pseudomonas aeruginosa* a human pathogen, which causes 10% of nosocomial infections.[148]

To guide future multiomic experiments, a comparative study is needed that directly relates and details the strengths and weaknesses of each method. Here a CME multi-omic extraction as well as an MTBE multi-omic extraction were performed to directly compare these methods for multiomic analyses. Additionally, the lysis technique was also examined, testing a CME sonication (CME-S) method and a CME freeze-thaw (CME-F) method to understand the effects of lysis techniques on biomolecule extraction fidelity. This sets up two binary comparisons where MTBE can be easily compared to CME-S and CME-S can be compared to CME-F. Unfortunately, due to the COVID-19 lab shutdown during the latter phase of these experiments, we were not able to as thoroughly validate the LC-MS data as desired. However, the goal is not to perform a comparative discovery metabolomics experiment, but rather to test multiomics extraction procedures for viability. Simply testing if lipids, proteins, and metabolites can be simultaneously extracted does not require optimized analysis methods. Similarly, some rough comparisons should be able to be performed between the three methods despite less than ideal replication and optimization, and further experiments can validate conclusions.

Material and Methods

Multiomic Extraction Procedures

Three different extraction procedures were tested. The first was a chloroform methanol-based extraction with sonication (CME-S) performed similarly to methods common in the field.[47, 51, 61] A cell pellet was resuspended in 1mL of PBS by vortexing. Two 500 µL aliqouts of this suspension were taken as two aliquots and centrifuged at 10,000 × g for approximately 3 minutes to produce two samples with identical biological composition. The supernatant was discarded and to each pellet was added 140 µL of methanol and of water. The cells were suspended and lysed by tip ultrasonication for 5 minutes (with a 10 seconds on, 10 seconds off cycle for a total time of 10 minutes). Following lysis, 280 µL of chloroform were added, the mixture was vortexed and placed on ice in an orbital shaker for 1 hour and 20 minutes to extract

and partition analytes. The tubes were then centrifuged at 10,000 × g for 15 minutes to induce phase separation and the fractions were collected (the aqueous top layer, organic bottom layer, and insoluble material (proteins) in the center.) The water layer regained turbidity over time, and in certain cases additional centrifugation was performed before LC-MS analysis.

A variant of this was also performed with freeze-thaw lysis instead of sonication (CME-F). In this case, cells were split and collected as in CME-S, but chloroform and methanol were added with volumes of 280 μL and 140 μL respectively. After resuspension via vortexation, these were then flash frozen in liquid nitrogen and allowed to thaw on ice. After thawing, tubes were vortexed to disperse the cellular material. This was performed twice more for a total of three cycles. Following freeze-thaw lysis, 140 μL of water were added, the solution was vortexed, and placed in ice on an orbital shaker to extract and partition. The extracts were then centrifuged and collected according to the CME-S method.

A third method was also tested using methyl-tert-butyl ether (MTBE) instead of chloroform. This is based on the Matyash et al. lipid extraction procedure, which has also been used for multiomic extractions.[60, 62] This was performed identically to the CME-S method except 125 μL of water, 150 μL of methanol and 500 μL of MTBE were added in the place of chloroform. Since MTBE is less dense than water, the organic phase was at the top and the insoluble protein was at the bottom of the vial. The MTBE fractions were dried under nitrogen to about half initial volume.

Crude Protein Analysis

The mass of the protein in the precipitated cell debris was measured using a Thermo Scientific Nanodrop spectrometer. The insoluble pellet was suspended through vortexing and was diluted by a factor of twenty before analysis with the spectrometer. The preprogrammed protein A205 Scopes method was used, which measures the protein concentration by measuring absorbance at 205 nm. The peptide bonds absorb strongly in the deep-UV, and this gives a method of quantification through the use of Beer's law. Measurements were taken on three different dilutions and averaged to find the final protein concentration.

Liquid Chromatography-Mass Spectrometry Analysis

Metabolomic and lipidomic analyses were performed in technical duplicate using an LTQ (linear trapping quadrupole)-Orbitrap Velos Pro instrument coupled to an Ultimate 3000 LC system. The LC was operated in a split-flow nano LC method and interfaced with the Velos using a nano ESI source. Columns had an inner diameter of 100 μm and were packed manually using a pressure cell to a length of 15 cm using 5 μm ZIC-pHILIC (zwitterionic-polymer) particles. HILIC analyses of both metabolites and lipids were operated with a gradient of 100% B to 40% B over 20 min, followed by a return to the initial conditions and a wash/re-equilibration for 15 min. Solvent B was 97% acetonitrile and 3% water with 5 mM ammonium acetate while solvent A was 100% water with 5 mM ammonium acetate. The use of HILIC avoids the issues RP has of extreme lipid retention and no retention of metabolites. The mass spectrometer was operated in a DDA mode with one orbitrap full scan followed by ten LTQ MS^2 fragmentation scans using CID at 35 normalized collision energy. For the lipid analyses, the scans were from 200 to 2000 m/z while the metabolite analyses were run from 50 to 1700 m/z. The reason for this mass range difference is due to the larger sizes typical of lipids as compared to metabolites.

Data Analysis

The data was analyzed using open source software mzMine.[149] Features were detected from the raw data using MS/MS peaklist builder (with a 10 ppm tolerance), deisotoped, duplicate peaks filtered out, aligned in retention time, filtered according to their duration (0.1-8 min) and height (2e3 intensity), and finally were gap filled. After this processing, features were searched against online databases with a 10 ppm tolerance. Lipid Maps was used for the organic data, and Metacyc as well as MassBank and KEGG was used for the aqueous data.[93-96] The samples were blanked by subtracting the blank peak area from that of its corresponding samples. Following this, all features that were at the background (≤0) for more than one sample in each set were removed.

Results

Lipids

Lipid samples (organic fractions) were run on the HILIC LC-MS method in both positive and negative mode. This ensured good coverage of different lipid classes which do not all ionize in the same mode. Phospholipids are polar enough to retain and be separated on HILIC columns. Non-polar/neutral lipids will most likely elute in the dead-volume, but these are not expected to ionize well in the ESI setup used. In this experiment, we focus on phospholipids which should be the most abundant and well represented by the experimental setup.

The lipidomics base peak chromatograms (BPCs) for each extraction method are displayed to demonstrate the presence of analytes in the organic fraction for each extraction (Figure 3-1). Across all extractions, the positive mode runs showed more than an order of magnitude greater signal than the negative mode runs. Despite this, peak shapes are generally similar between modes. MTBE had the richest BPC with much of the signal spread out over 5 to 10 min. The CME BPCs had different shapes than MTBE, though similar to each other. These had two peaks, one centered at 9 min and one at 14 min. Despite different signal intensities, a similar number of features were detected by mzMine. There were about one thousand features for each sample set in each mode. MTBE showed significant differences from its analogue CME-S in that more than half of the features detected in that pairing were specific to only one condition. CME-S and CME-F had much in common and slightly more features were detected with CME-F.

The features were annotated by matching to the lipids maps database (searched with 10 ppm mass accuracy) and the ID results are summarized in Table 3-1 and Figure 3-2. The total carbons for lipid tails range from 13 to 44 with degrees of unsaturation up to 8, with 0-2 being the most common.

Phosphatidylserine (PS) was the lipid class with the greatest number of IDs. These were more commonly annotated in positive mode than in negative mode and range from 81 to 92 in positive mode and 67 to 77 in negative mode. Additionally, the peak area is 23-fold higher on average in positive mode compared to negative mode.

The second most frequent class ID was phosphatidylcholine (PC) with up to 67 feature IDs. More PC was detected in positive mode than negative mode for MTBE, but more in

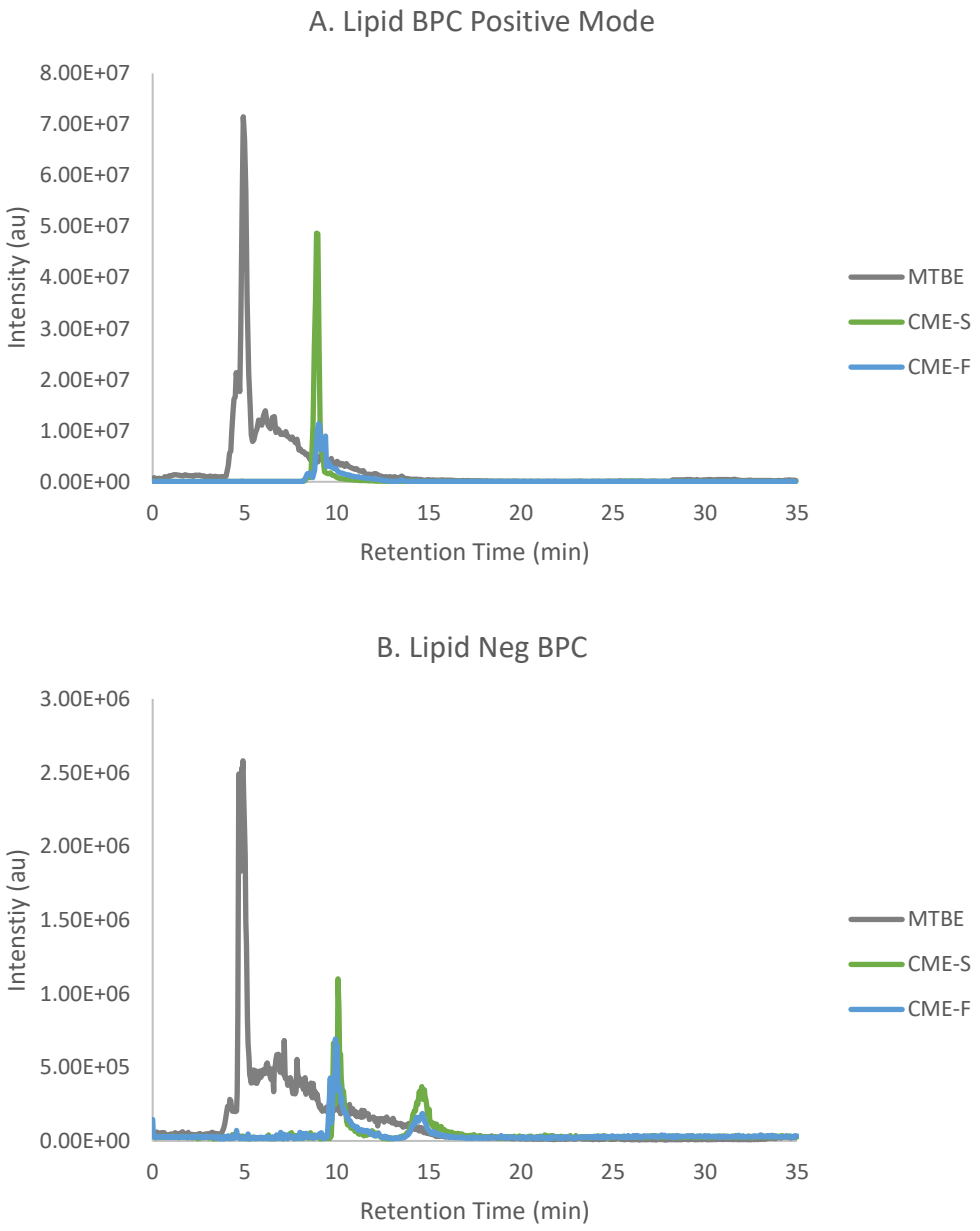

Figure 3-1. Lipidomic Base Peak Chromatograms. This BPC are shown for the organic fraction of each extraction in A. positive mode and B. negative mode.

Table 3-1. Phospholipid Annotations by Sample. Each tentative annotation of a standard phospholipid is given below.

Lipid Species	CME-F Positive	CME-F Negative	CME-S Positive	CME-S Negative	MTBE Positive	MTBE Negative
CL	30	12	22	9	16	16
PA	12	24	10	22	16	26
PC	62	65	56	55	67	60
PE	2	5	1	5	4	5
PG	6	46	6	40	6	46
PI	5	5	2	5	5	6
PS	90	72	81	67	92	77

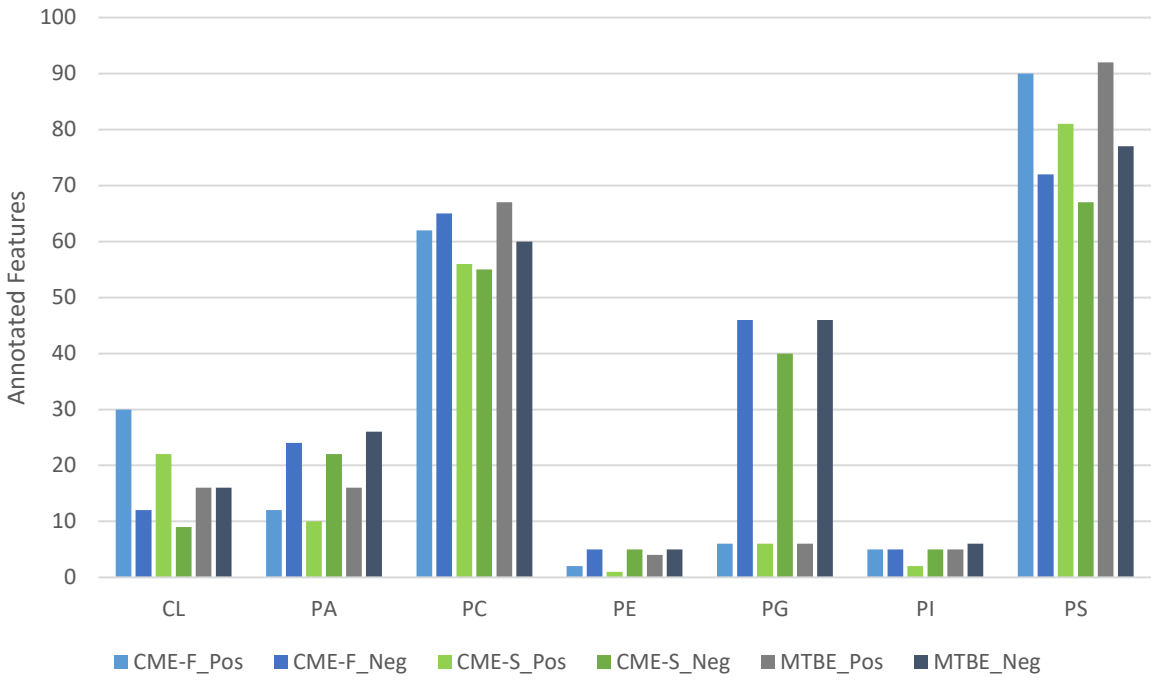

Figure 3-2. Lipid Class Annotation Bar Graph. Number of annotated features belonging to each class in each fraction.

negative mode than positive mode for both CME methods. Despite this, these PC lipids were well over an order of magnitude more intense in positive mode than in negative mode.

Phosphatidylglycerol (PG) was detected with up to 46 annotations in negative mode, however very few PG lipids were annotated in positive mode (6 in each sample). An example of a PG negative mode feature is shown in Figure 3-3, with fragmentation allowing tail lengths to be decerned. Cardiolipin (CL) was annotated more in the CME extracts than MTBE with 30 in CME-F, 22 in CME-S, and 16 in MTBE (all positive mode). Phosphatidic acid (PA) had about 25 IDs in negative mode and 10 to 16 in positive mode with slightly more IDs in MTBE than in CME. There were very few phosphatidylethanolamine (PE) and phosphatidylinositol (PI) annotations with about 5 across the board for both.

A rough quantitative comparison of the method was performed by plotting a heatmap of the peak areas (relative to the mean) for all of the features with lipid annotations. This was done with the positive mode data as that gave the best signal and can be seen in Figure 3-4. The MTBE extraction gave greater yields overall, followed by CME-F and then CME-S.

Average relative standard deviations (RSDs) of peak areas for features present in each replicate of an extraction were calculated for technical replicates and compared to the overall relative standard deviation to evaluate reproducibility. The MTBE method had average RSDs of 46% for both sets of technical replicates and an overall RSD of 56%. CME-S gave technical replicate RSDs of 45% and 68%, but an overall deviation of 51%. CME-F gave technical replicate RSDs of 42% and 43%, but an overall deviation of 52%. These values were consistent with the negative mode data where MTBE gave technical RSDs of 48% and 56% and an overall RSD of 68%. CME-S gave technical RSDs of 49% and 44%, and an overall deviation of 49%. CME-F gave technical RSDs of 42% and 40% and an overall deviation of 51%.

Proteins

After protein quantification by nanodrop assay, the following data were obtained. The MTBE extracts gave 2.15 and 2.07 mg of total protein (a yield previously found to be approximately 10% of the starting cell pellet) while the CME-S gave 1.62 and 1.55 mg, and the CME-F gave 1.48 and 2.60 mg. It is of note that the different extraction methods gave qualitatively different protein precipitates. CME-S and MTBE gave protein precipitates that were

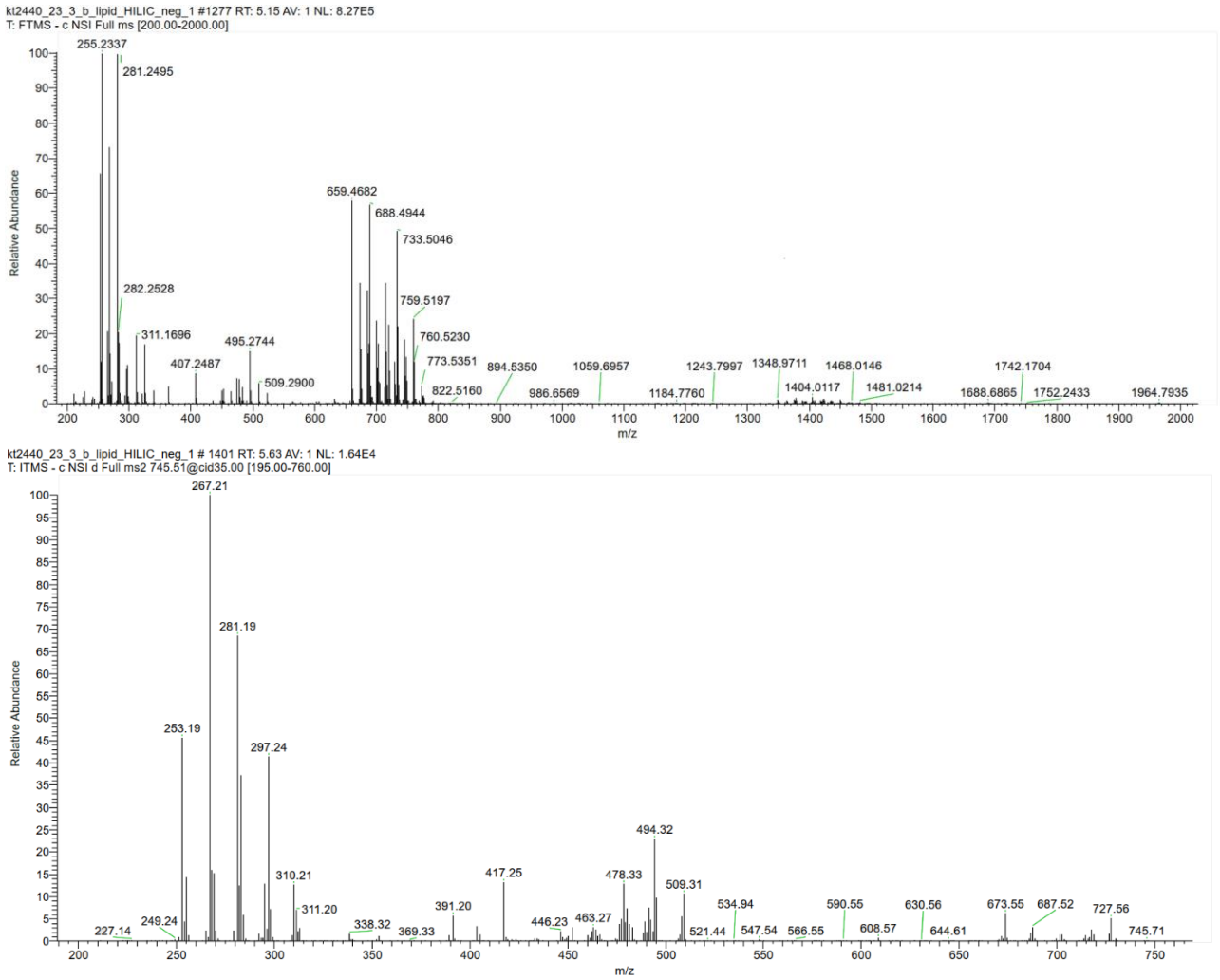

Figure 3-3. Mass Spectrum of PG (34:2). The mass spectra for a Phosphatidyl glycerol (34:2) feature are given. The full scan (A) and tandem MS are shown (B). The parent mass is 745.51 m/z. The 267.21 and 478.33 peaks indicate the presence of an 18:1 tail, and the 253.19 suggests the presence of a 16:1 tail, supporting the 34:2 assignment.

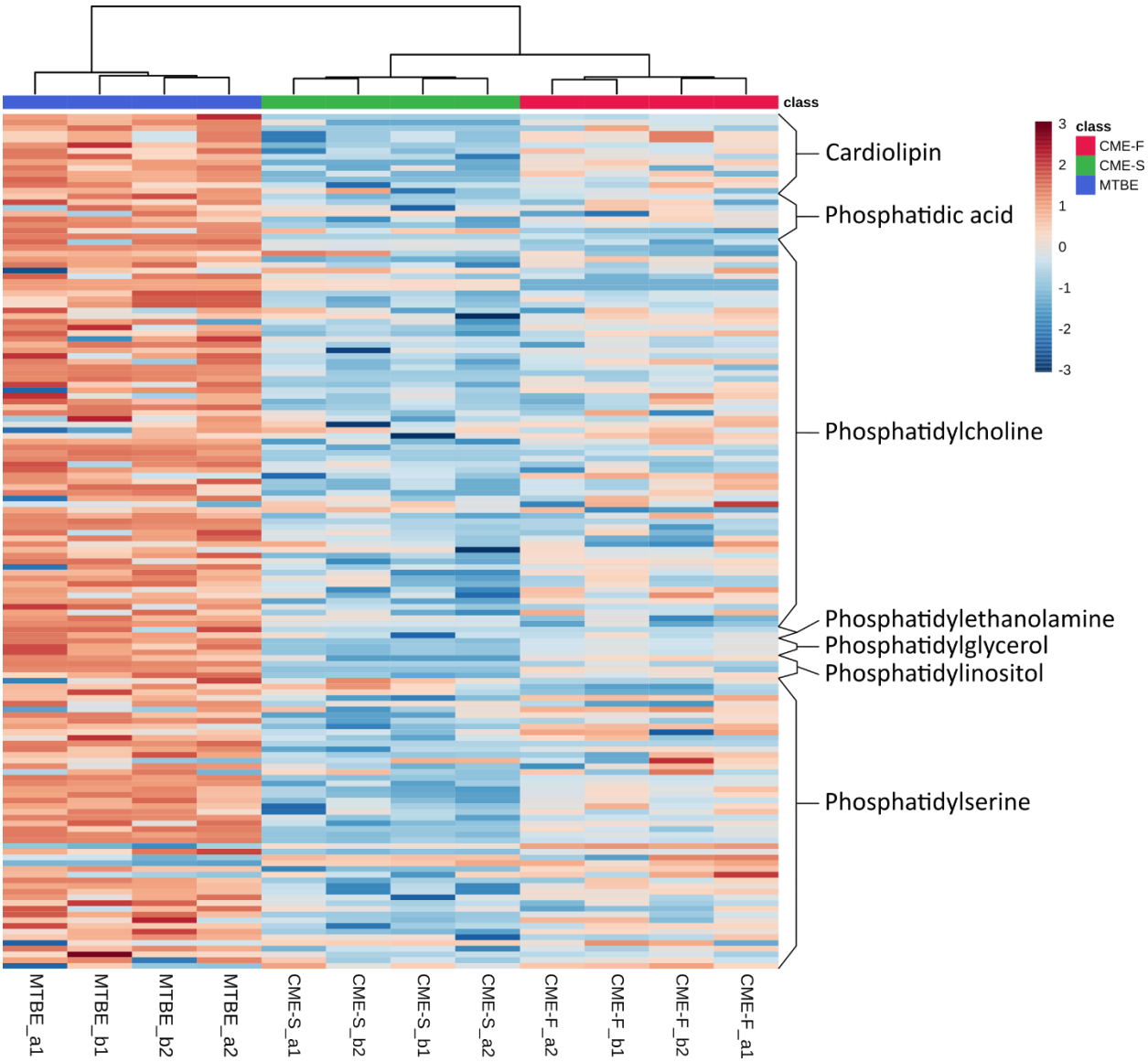

Figure 3-4. Lipid Heatmap. This displays the foldchange of the log-transformed peak areas for each lipid annotated feature relative to the mean. Generated using Metaboanalyst.[99]

white and had a somewhat dry appearance while the CME-F was off-white or tan and had something of a slimy appearance.

Metabolites

The metabolite (aqueous) fractions were analyzed according to the same LC-MS procedure as the lipids except the preliminary negative mode data were poor and so positive mode was focused on for this study. Although RP may be included in future studies, HILIC is expected to perform better for the polar metabolites, as many metabolites will not be retained by RP and elute in the dead-volume. An ion-pairing reagent is disfavored from use which would improve retention but impede ionization, as both positive and negative mode data is desired (at least for the lipids). The BPCs can be seen in Figure 3-5, and they were all fairly comparable with small peaks between five and eighteen minutes and a large peak centered at fifteen minutes. The peak at fifteen minutes is primarily due to sodium acetate clusters.

After processing data using mzMine with a 10 ppm mass tolerance, more than one thousand features were detected, the results of which are summarized in Figure 3-6. There are few differences in the features detected between the CME methods. However, the MTBE method and the CME-S method show a surprising amount of diversity in the features detected where more than half of the features are specific to an extraction method with more features detected in MTBE.

Annotations of the features include a variety of different metabolites including amino acids, nitrogenous bases, and nitrogen metabolites, though there is a lack of central carbon metabolites. Examples include alanine, cystine, guanine, and urea. Many energy molecules were annotated, with many nucleotide phosphates throughout. These bear further consideration as they give a clue into the fidelity with which the extraction represents the metabolic state. Degradation is especially an issue for metabolomics. Limiting metabolite degradation is critical to capture an accurate picture of the cellular physiological state, but many of these are labile and can easily degrade, either abiotically, or through enzymatic action. The nucleotide phosphates are some of the most notoriously unstable metabolites and are the best choice to assess degradation. Those

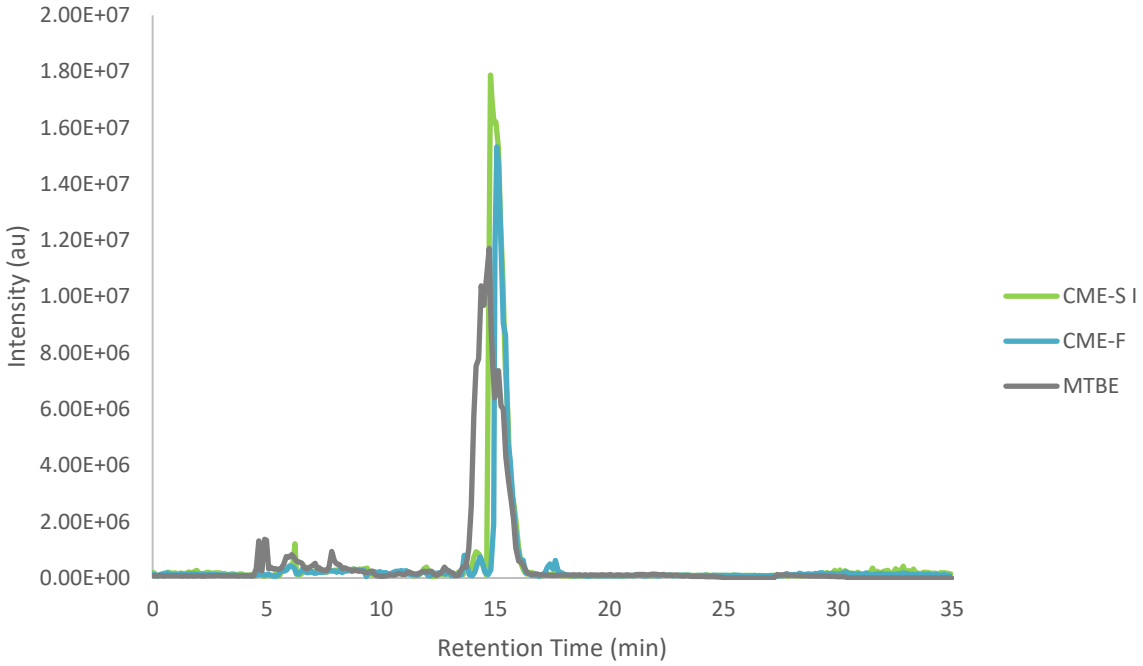

Figure 3-5. Metabolomics Base Peak Chromatogram. This displays the BPC for each aqueous fraction.

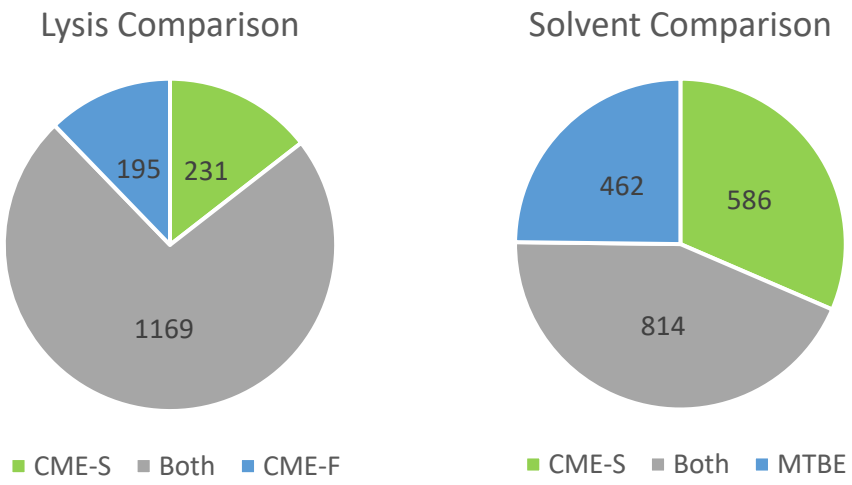

Figure 3-6. Shared Metabolite Features in Extracts. This pie-chart breaks down the total features based on by similarity across extraction procedures.

annotated in the feature list include ATP, CTP, CDP, CMP, GTP, GDP, GMP, UTP, UDP, and UMP. After further examination, the GXP group of nucleotide phosphates were shown to give the most reliable data, and so these were used to compare the extraction methods for degradation potential. Figure 3-7 demonstrates the GXP EICs for each set of samples. Note that the retention times increased slightly when going from GMP to GDP to GTP, as is consistent with increased hydrophilicity. GTP was not reproducibly detected in the CME-S extract (2 of 4 runs), but of those two runs, the GTP/GDP ratio was 4.09. For CME-F and MTBE extracts the ratios were 6.35 and 3.51 respectively. GDP/GMP ratios followed a similar pattern with values of 0.339, 1.88, and 1.26 for CME-S, CME-F, and MTBE respectively.

To provide a quantitative comparison of the feature intensities, a heatmap was generated comparing the foldchange of every feature in each sample relative to the mean. This can be seen in Figure 3-8. Blocks of features associated with different extraction methods, but none had clearly higher intensities globally.

The reproducibility of the extraction was again estimated by comparing the RSDs of the technical replicates with the overall RSD. The MTBE extraction had technical RSDs of 39% and 36% with an overall RSD of 45%. The CME-S extraction had technical RSDs of 36% and 39% with an overall RSD of 49%. The CME-F extraction had technical RSDs of 32% and 37% with an overall RSD of 42%.

To determine how much lipid content was present in the metabolite fraction, the data were run in mzMine against Lipid Maps as the lipid fractions were. A moderate number of phospholipids were detected in the metabolomics samples with 1 PA, 20 PC, 4 PG, and 7 PS for the MTBE method. For the CME-S method, there was 0 PA, 13 PC, 1 PG, and 2 PS.

Discussion

Lipids

The BPC for the MTBE fraction demonstrated good signal (at least in positive mode) and decent separation with signal spread across an eight-minute period (Figure 3-1). This is promising, demonstrating a good amount of analyte in the extract. The CME BPCs, on the other hand, show mostly a clump at 9 minutes (and at 14 min. for negative mode). Although there is still good signal, this means most analytes will be eluting at the same time, which may will hinder analysis through ion-suppression. Additionally, if too many analytes elute at the same

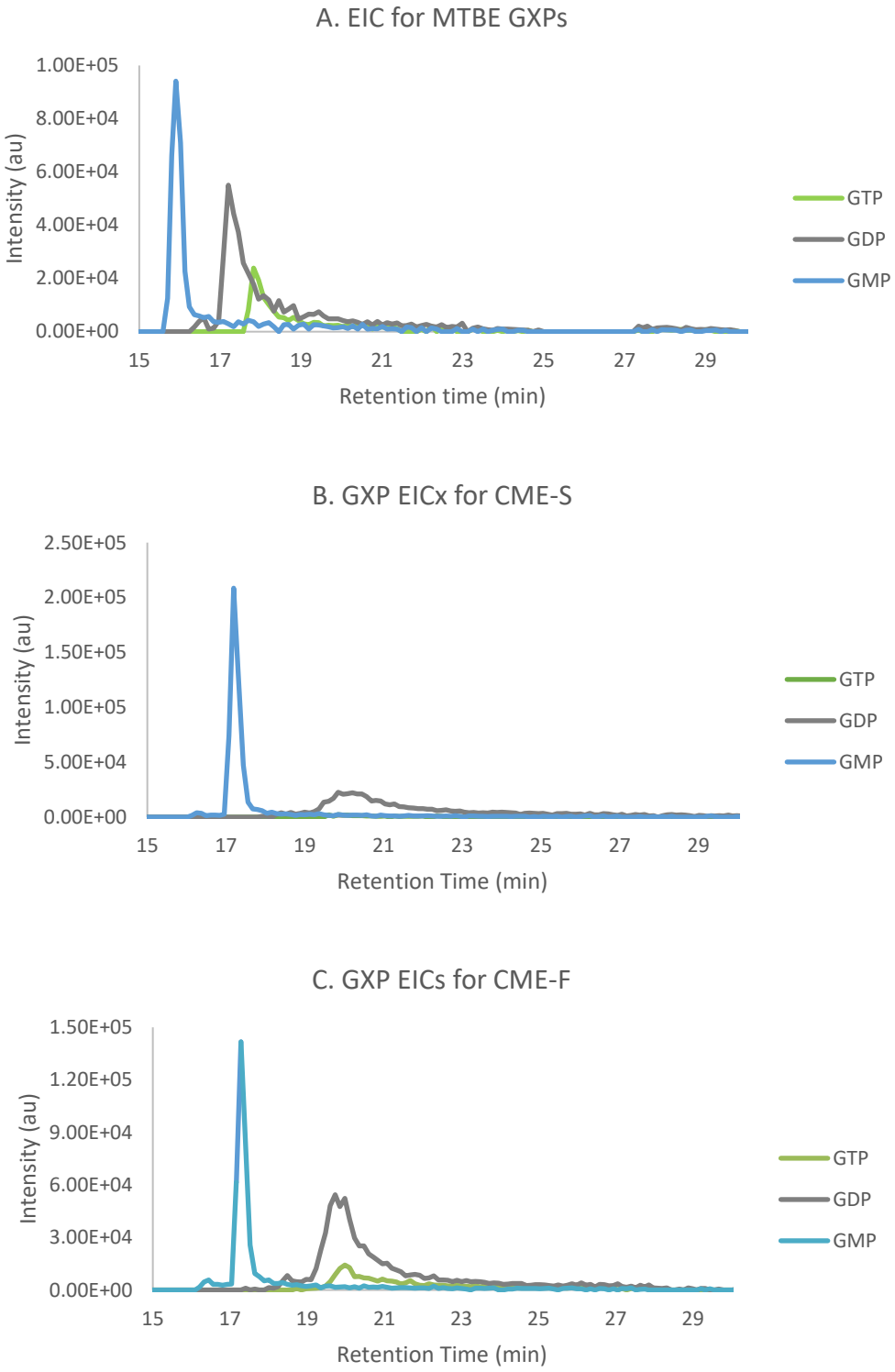

Figure 3-7. Extraction Ion Chromatograms of GXPs. The EICs for guanosine triphosphate (GTP), guanosine diphosphate (GDP), and guanosine monophosphate (GMP) are given for each extraction condition, A. MTBE, B. CME-Sonication, CME-Freeze thaw.

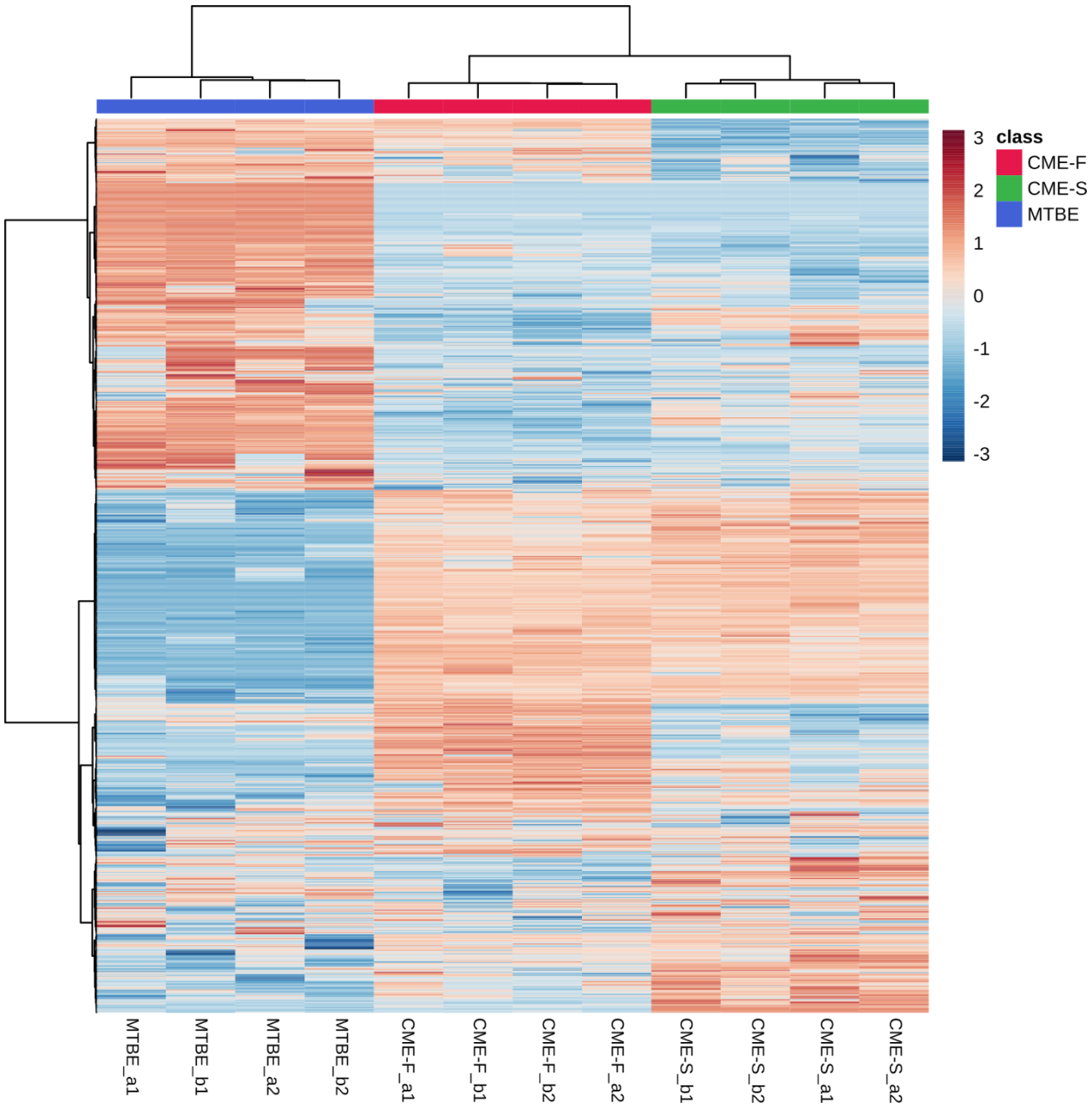

Figure 3-8. Metabolite Feature Heatmap. This displays the foldchange of the log-transformed peak areas for each metabolite feature relative to the mean. Generated using Metaboanalyst.[99]

time, the DDA cycles may not be fast enough to select all analytes for fragmentation. The signal drop seen in both the BPC and the average peak areas from positive to negative mode is intriguing and could be due to differential ionization efficiencies of the lipids. Overall, the BPCs are promising and display what seems to be a successful extraction of lipids.

Looking at the annotations may reveal more about the extraction efficacy. PS was the most annotated lipid, with large numbers of the species detected. The PS annotations demonstrate an ionization preference for positive mode based on peak area. This is a zwitterionic lipid and should ionize in both modes, so it is odd to see such a large increase in the peak area in positive mode. This could imply some technical difficulties with negative mode, such as a non-optimized spray voltage. Regardless, large numbers are seen in both modes. This indicates PS as a major contributor to the lipidome; however, the presence of PS in such high numbers is surprising as PS is thought to be primarily an intermediate to PE and not a major membrane lipid in *P. putida*.[150] While some of these extracted ion chromatograms are poor, others show decent peak shapes (Figure 2-9), so it appears these are real analytes. Since these are just tentative annotations, more verification of the IDs (including the use of standards) would help to shed light on the abundance of these components and the accuracy of these IDs.

PC, with the second most annotations, is also a surprising ID for two reasons. First, large amounts are annotated in negative mode. Due to the quaternary ammonium in the PC headgroup, an M–H PC species carries no charge. The lack of exchangeable protons excludes the possibility

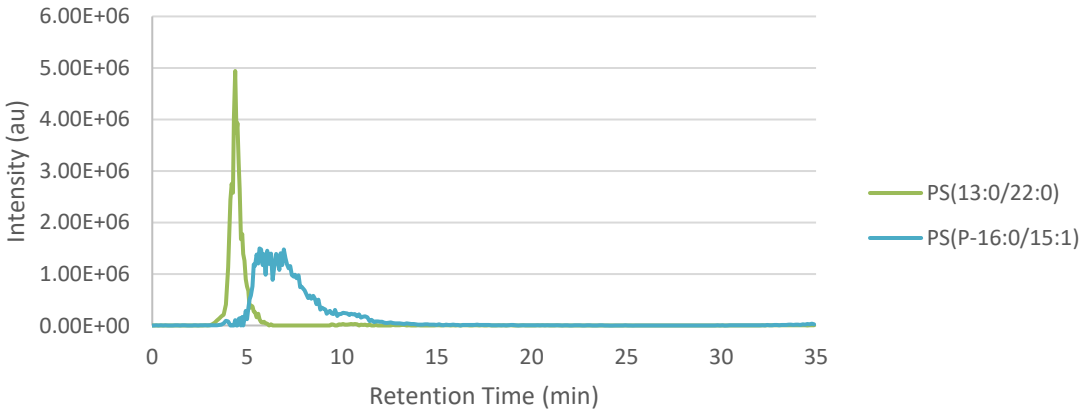

Figure 3-9. Extraction Ion Chromatograms for Two PS Lipids. This shows the EICs for two PS lipids, PS (13:0/22:0) and PS(P-16:0/15:1)

of M–2H ions in ESI, making ionization in negative ESI impossible without adducts, such as chloride. The database was not searched for adducts other than hydrogen (M±H), so the PC annotations in negative mode seem to be impossible. One explanation of this is that these are tentative annotations and could simply be incorrect. Another explanation exists, however, that these are not false, but rather mislabeled. PE differs from PC by only C_3H_6, which means a PE with n saturated carbons and a PC with n – 3 have identical masses. An examination of the literature shows that PC lipids are not believed to exist in *P. putida* KT2440, while PE is one of the more abundant phospholipids.[151] Given that there are extremely few direct PE annotations in the data, mzMine seems to favor PC annotation even when these are nonsensical. Thus, these PC annotations are more likely to be PE with a net tail length three carbons greater. This assignment fits the data best, as PE can ionize in negative mode, but is more suited to positive mode. This helps explain the discrepancy between the BPC signal intensities as well as the increased peak area of the PC (PE) features in positive mode and its presence in negative mode.

PS and PE, based on annotations, seem to be the largest contributors to the *P. putida* lipidome. Phosphatidylglycerol is the next greatest contributor. This is one of the main lipid classes expected to be present in *P. putida*.[151] Based on annotations, this shows a strong preference for negative mode—which is consistent with the chemistry of the molecule. The intensity is low in general for these features. This may be due to poor ionization or just low abundance of PG lipids.

Cardiolipin and phosphatidic acid also have minor contributions to the lipidome, while phosphatidylinositol seems to be negligible. Phosphatidic acid is also not expected to be a major membrane contributor and could be present merely as an intermediate in the synthesis of other phospholipids.[151] Cardiolipin has been reported in *P. putida* before and is not a surprising ID.[151] The most cardiolipin IDs are in the CME-F (positive) data set followed by CME-S and MTBE. This may be due to the fact cardiolipins are larger lipids and more hydrophobic and chloroform, being the less polar solvent, may capture these better.[152]

Overall, the three major lipid classes in *P. putida*, PE, PG, and CL, are observed, implying the extractions tested are viable options for bacterial lipidomics. More examination is necessary to determine the source of the large amount of unexpected PS lipids. Slight differences in the lipid profiles between MTBE and the CME methods were observed, but not enough to explain the large differences in the BPC profiles. Technical difficulties ought also to be

considered as they may play a role in the observed differences. The CME extracts were run at a later point with a different column than the MTBE samples. The column, although packed to the same length, had an extended section of unpacked tubing. This could have resulted in a retention time shift, as well as compacting the range of elution, which would explain the differences in the BPC. The intensity differences between ionization modes seem to be due in part to differential ionization of the lipids, but it could also be due to a lack of optimization of the negative mode ESI conditions. Perhaps the HPLC solvent additive could be altered to give superior ionization in negative mode.

Comparing the two CME methods against each other, they had very similar BPCs for both modes with the exception that CME-S gives stronger signals than CME-F. This is to be expected if freeze-thaw does not lyse all cells, as the cell debris indicates, since intact cells will precipitate with the other cell debris and their lipids will be sequestered in the precipitate. However, after examining the feature annotations, CME-F consistently results in slightly higher total features, as well as annotations per class, than CME-S does. An examination of the peak areas of the heatmap in Figure 3-4 reveals that CME-F also gave greater average peak areas than CME-S does. Again, this is hard to resolve with the BPC, which shows greater signal for the CME-S than for the CME-F. Although degradation may be an attractive explanation because degradation would lower analyte peak areas while increasing those of degradation products, which may not be annotated by mzMine, this is not likely since degradation is uncommon for phospholipids. Perhaps there are components other than phospholipids that are contributing to the observed differences and are extracted more readily by the CME-S method. Further testing will be needed to verify this. Triacylglycerols (TAGs), however, were not detected and only a few diacylglycerols (DAGs) were detected (8 in positive more, 10 in negative), meaning DAGs and TAGs are not a significant part of these hypothetical other components.

Although the system was not validated for quantitative evaluations, a rough analysis of the reproducibilities of the peak areas was performed for each sample set by calculating the peak area RSDs. Since biological variation is controlled by extracting from the same cell pellet, the variations should be limited to that generated in the extraction and the LC-MS system. The RSDs of the technical replicates should include only that deviation caused from the LC-MS system, and by comparing to the overall RSD, the relative contribution of the extraction procedure can be evaluated. The RSD for MTBE positive data overall was 56% while that of the individual

replicates were 46%. This indicates that the extraction procedure adds about 0.1 or 10% to the RSD, a little more than a one-fifth increase. This is a moderate increase that is not excessive. CME-S exhibited little to no increase and CME-F also had about a 0.1 or 10% increase as well as MTBE. For the negative mode data, MTBE had an increase of about 0.16 or 16% and CME-S again had little to no increase while CME-F increased by about 0.1 or 10%. Although MTBE had a somewhat higher RSD, these are all in the same range as the positive mode RSDs, and given the signal drop across modes, it is surprising that they are not worse than the positive mode RSDs. The extractions, then, do not seem to be adding excessive uncertainty to these measurements.

Proteins

All extracts gave copious amounts of protein that are more than sufficient for proteomic analyses. The physical differences observed in the lysates is relevant, however. Because the CME-F pellets had a more slimy and off-white appearance, this implies that the cells were not fully lysed by the freeze-thaw method, which could be represented in its inconsistent protein yield. To better characterize the extraction viability for protein analysis, the recovered protein should be analyzed through proteomics. However, that was beyond the current scope of this work, and future experiments hope to elucidate the relationship between solvent, lysis, and protein yield. This test shows that these methods extract sufficient protein for proteomic analysis, making these viable for proteomics as well as lipidomics with the possible exception of CME-F. It is worth noting that coupling this protein collection to a proteomics sample preparation method would be trivial.

Metabolites

To determine the efficacy of the extraction for polar metabolites, the BPCs for the samples were first examined. The large peak in the BPC was identified to be sodium acetate clusters by a repeating 82.0032 *m/z* unit (sodium acetate's exact mass is about 82.0031 Da), see Figure 3-10. Acetate is present in the HPLC solvents, and the sodium is probably from residual PBS that was used to split the cell pellet. If a washing step is employed, or with harder centrifugation, the PBS could potentially be completely removed. However, washing may also lead to the leakage of metabolites. Beyond the large peak, there are smaller peaks in the

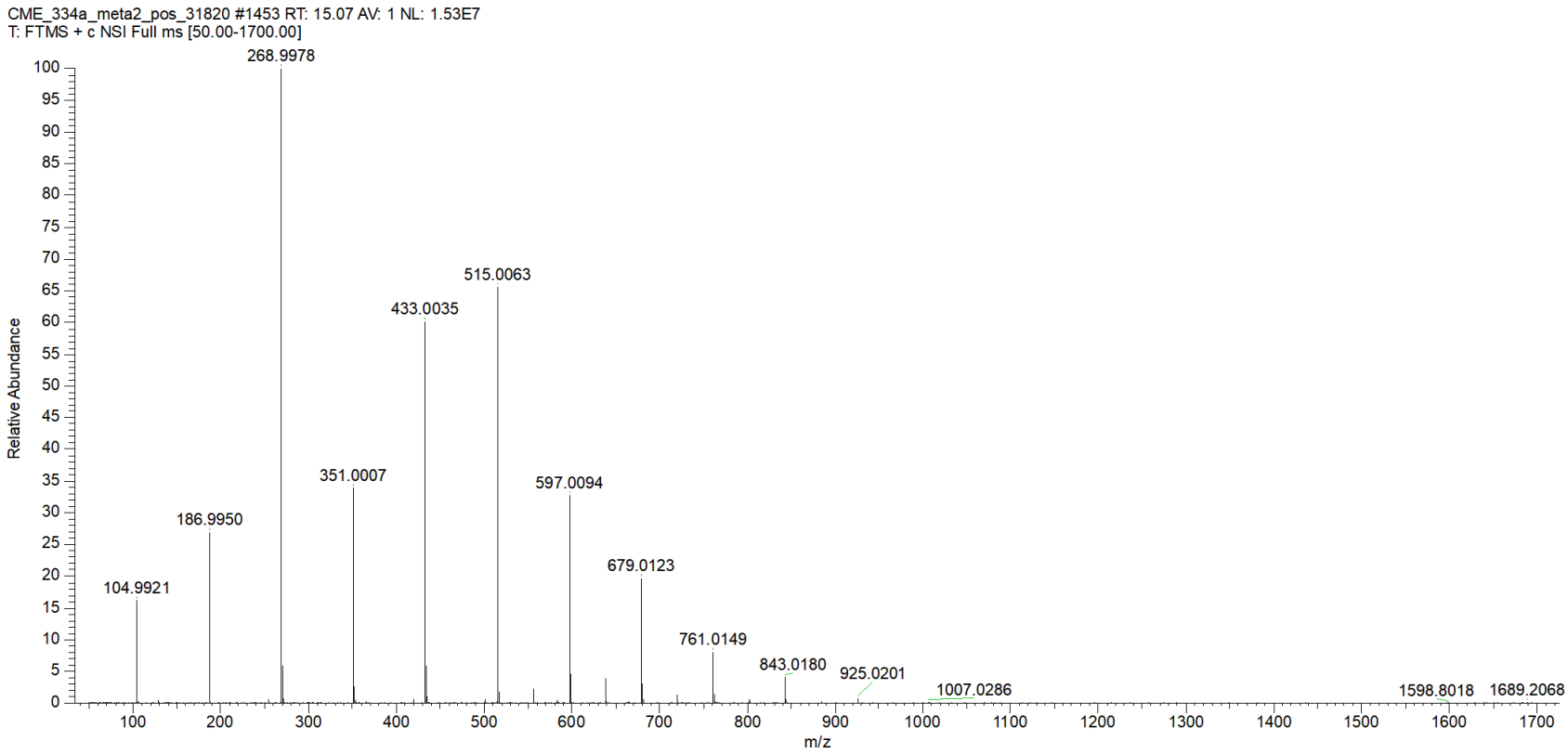

Figure 3-10. Spectrum Showing Sodium Acetate Adducts. This spectrum shows the repeating 82.0032 *m/z* peaks assigned to sodium acetate clusters.

chromatogram, which can be seen throughout. This indicates that the extractions are indeed successful in collecting metabolites. This is reinforced by the observation that mzMine annotates approximately one thousand features in each method, a rich enough data set. CME-S and CME-F have almost all of their features in common, whereas CME-S and MTBE share less than half of the unique features between the two. Potentially, this could be related to different selectivities of the solvents involved. The solvents were balanced such that an approximately equivalent amount of methanol and water were added in both the MTBE and the CME methods. So, differences in partitioning or extraction efficiency would result from the organic solvent. However, more, or different, features may not necessarily be a good thing. It may mean that analytes are spreading between both the organic and aqueous layers, thereby diluting the analytes and making the extracts unnecessarily complex. To help account for that, the annotated features were examined for commonalities with the organic fraction. MTBE had twice as many (16 more) phospholipid annotations after running through Lipid Maps as CME-S did. Some of these features may be related to lipids that do not fully partition into the organic layer. This implies the chloroform has superior performance in extracting lipids into the organic phase relative to MTBE. This performance difference is likely greater than it first appears, as there was approximately twice asmuch MTBE used in the extraction as chloroform, which ought to favor MTBE drawing more lipids out of the aqueous fraction. This differential extraction may be caused by MTBE's greater polarity not extracting the lipids as efficiently as the chloroform-based solvent.[152] To better characterize different extraction profiles, van Krevelen diagrams were generated using the Open van Krevelen software (Figure 3-11).[103] These diagrams were generated with a 10 ppm tolerance and ions below 15% of the most intense ion were filtered to clarify the analysis. Based on the van Krevelen profiles, the extractions appear to have similar compositions.

One factor that was particularly intriguing is whether the lysis method played a role in metabolite degradation. To assess this, the ratio of energy metabolites was calculated. The GXPs were chosen to highlight this comparison as they had the best quality data. That can be seen from the EIC as well as the accompanying spectra, a Full-Scan showing GMP as the most intense peak and a tandem MS showing a strong 152 *m/z* peak that fits a guanine fragment (see Figure 3-12). GTP/GDP and GDP/GMP ratios are good way to assess degradation because degradation will tend to drive those ratios down. Therefore, in general, the higher the ratios, the less degradation. Ratios of intracellular GTP/GDP vary depending on the source, but taking an average from three

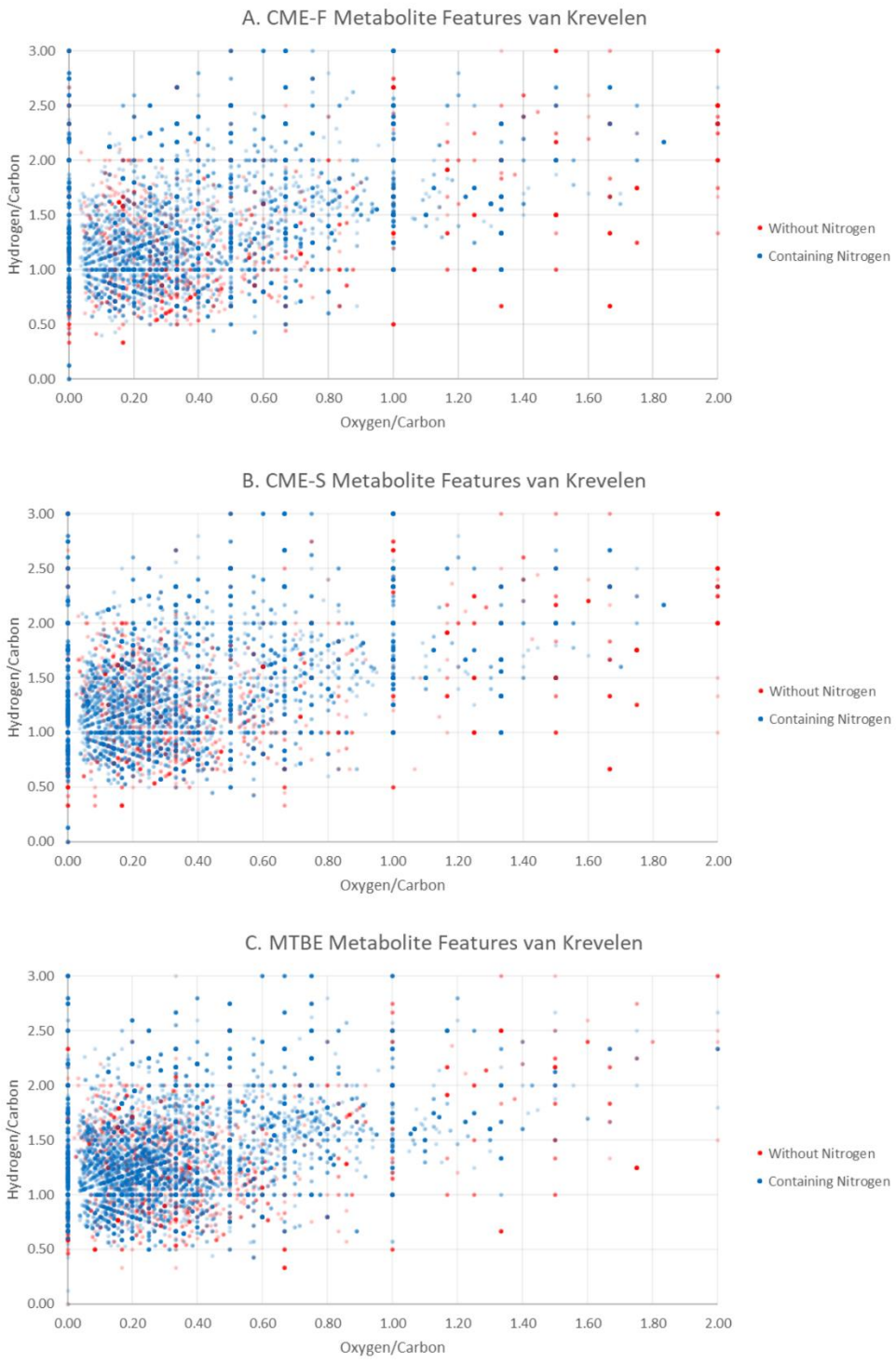

Figure 3-11. van Krevelen Diagrams of Metabolomics Results. van Krevelen diagrams are shown for each of the extraction types. The x-axis plots the oxygen/carbon ratio and the y-axis plots the hydrogen/carbon ration. A. corresponds to CME-F data, B. corresponds to CME-S data, and C. corresponds to MTBE data.

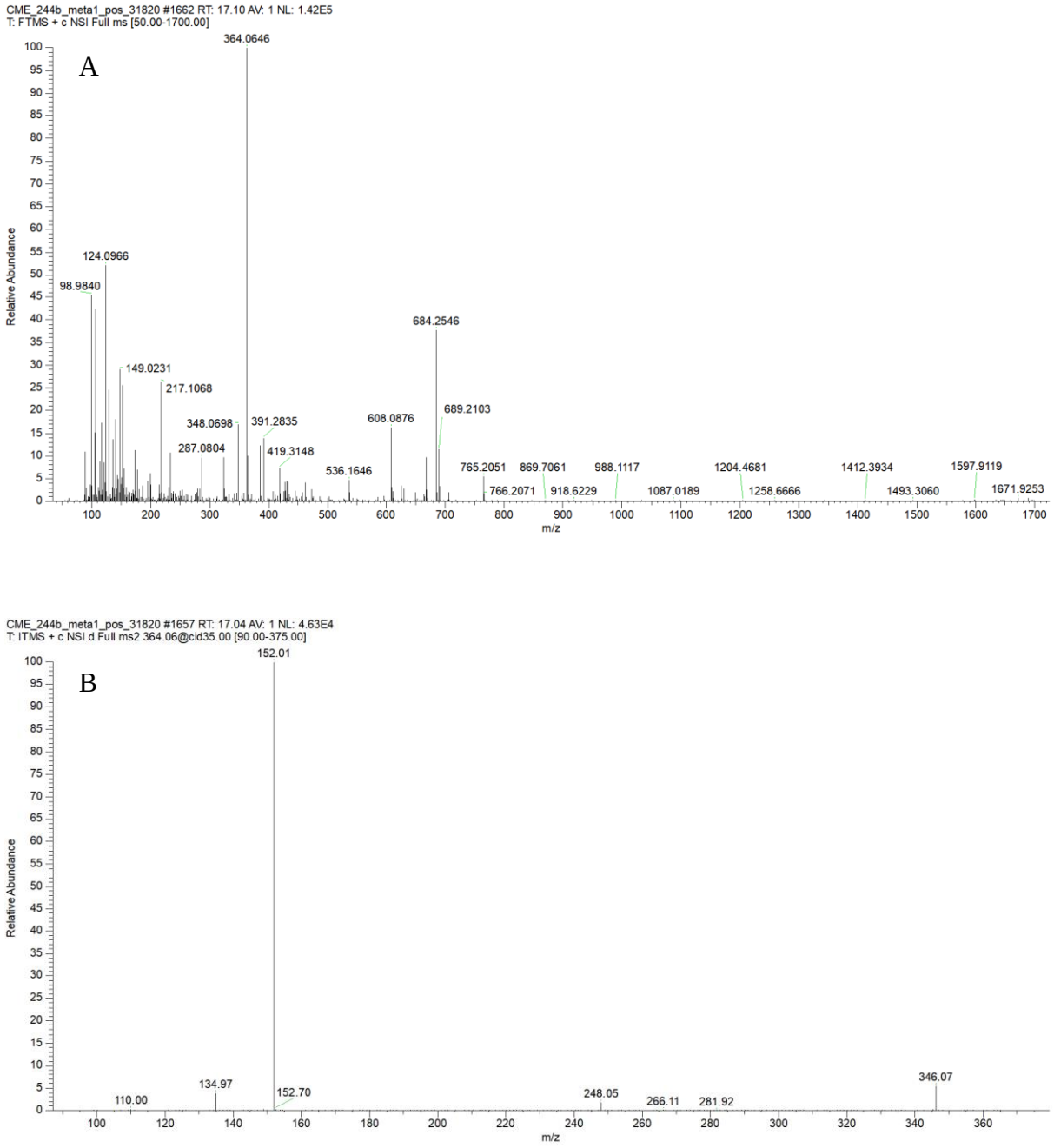

Figure 3-12. Spectra for Guanosine Monophosphate. The spectra for GMP in the CME-S extract is shown. A shows the full-scan where CMP is the base peak and B shows the MS^2 scan taken with GMP selected for fragmentation.

different sources gives a value of 7.5.[35, 153-154] This is moderately close to the value obtained for the freeze-thaw method (about a 15% difference). CME-S was different from the literature value by 46% and MTBE by 53%. The GDP/GTP ratio averaged from two sources is 2.92 this gives percent differences of 36%, 88%, and 57% for CME-F, CME-S, and MTBE respectively.[35, 154] The ratios demonstrate that CME-F preserves high energy molecules the best with a low percent error relative to a literature average (and the ratio actually exceeded that of one of the citations).[35] Additionally, MTBE may perform slightly better than CME-S as GTP was not reproducibly detected in CME-S and the GDP/GMP ratio was higher for MTBE. It is not surprising that the sonication lysis would perform worse than the freeze-thaw, as during sonication the sample sits at room temperature for a lengthy preprocess, which can itself cause significant local heating. With the freeze-thaw method, the solution stays at or below 0 °C throughout the lysis process. Although unclear, the prospect of MTBE limiting degradation more than CME-S is surprising, and it would be very interesting to test an MTBE-freeze-thaw extraction to more conclusively determine if the solvent plays a role in metabolite degradation. It is worth noting, that it is possible these differences could also be due to changes in the extraction efficiency and not necessarily degradation.

The reproducibility was again not greatly affected by the extraction procedure with increases in RSDs of about 7.5% for MTBE, 11.5% for CME-S, and 7.5% for CME-S. The MTBE method is again more similar to the CME-F than the CME-S although the procedure is more similar CME-S. This could mean that both the use of MTBE over chloroform and the use of a freeze-thaw lysis over sonication individually improve results. A freeze-thaw lysis with MTBE solvent would be an illuminating companion experiment to investigate what solvent effects may be present.

Conclusions and Future Directions

Returning to the fundamental goal of this experiment, are these extractions successful for the simultaneous extraction of lipids, proteins, and metabolites? Despite ways to improve the analysis, the answer appears to be yes. Lipid fractions show good BPC signals and rich numbers of features, protein analysis reveals abundant protein yields, and although the metabolite analysis leaves room for improvement in terms of overall signal and certain IDs, identifications do

include important and challenging metabolites such as high energy nucleotide phosphates. This corroborates the previous evidence in the literature about the effectiveness of these extraction methods.[60-61]

The second goal was to compare the different solvents and lysis methods against each other to test differential extraction efficiencies. Overall, these three different methods behave fairly similarly while the CME methods may give cleaner extracts based on lipid presence in the aqueous fraction. CME-F yields the best GTP/GDP and GDP/GMP ratios, with MTBE slightly better than CME-S. MTBE also saw larger peak areas for lipidomics following by CME-F, then by CME-S. suggesting the best method may in fact be a MTBE freeze-thaw method; however, particularly for proteomics, more tests should be done to determine if inadequate lysis with freeze-thaw is an issue. One other factor to consider is the convenience between chloroform and MTBE. Although similar methods, MTBE simplifies fraction collection slightly by pelleting the proteins on the bottom of the centrifuge vial. Additionally, MTBE is not a probable carcinogen as opposed to chloroform, which is a marked improvement from a safety perspective. Therefore, although further experimentation is needed to validate findings, based on these preliminary results, the MTBE extraction procedure is recommended out of the three tested for multiomic extractions.

There are various avenues for future research to improve on this study and more rigorously validate observations. Although samples were run with duplicate extractions and duplicate technical replicates, for a total of four runs per extraction, this is a relatively small sample size and performing the same experiment with triple or quintuple replicates would be an easy way to generate more meaningful data. As previously mentioned, it is unclear if the PS annotations are accurate as these are surprising to find in a *P. putida* extract. Using PS standards and probing the tandem mass spectral data more deeply should help to clarify this. Also, lipid features can be investigated more fully for components other than phospholipids, which may help clarify the source of the signal discrepancies between CME-F and CME-S. Further optimization of the LC-MS method will also add needed rigor to the comparisons particularly the quantitative ones such as the GXP ratios and heatmaps. Finally, the database annotations by mzMine deserve further investigation. There are several features with poor extracted ion chromatograms, calling into question the existence of those features. Also, multiple features will, at times, share identifications, an issue that, again, could be solved by standards and further MS^2

investigation. Despite definite room for improvement, this is a foundational experiment that can be built upon to fully characterize these extractions and more easily integrate metabolomics in multiomics experiments.

Perspectives

In chapter 1, metabolomics was discussed and put into context with other omic techniques. The various challenges of other omics and the incomplete physiological picture they give were discussed as reasons for pursuing metabolomics. Metabolomics is open to the same criticisms, however. The experimental protocols are not well established and harder to perform than proteomics or genomics. The informatic backend is even further behind the experimental side. Biological interpretation is often harder than other omics despite being closer to the phenotype (compared to genomics or transcriptomics), as separating signal from noise, artifact from critical feature is extremely challenging. Additionally, there is not always a good way to fit together metabolomics results to form a clear biological picture. In chapters 2 and 3, however, metabolomics has been described as it fits in relation to other omics experiments. By stacking the analyses, outliers can be sorted out, physiological differences can be exposed, and observations confirmed or corrected. This is the future of metabolomics. This is the future of all omics and systems biology, the integration of multiple complex analyses to yield a web of data that can be teased apart to elucidate the chemical underpinnings of life.

References

1. Kivirikko, K. I., Chapter 9 Posttranslational processing of collagens. In *Principles of Medical Biology*, Bittar, E. E.; Bittar, N., Eds. Elsevier: 1996; Vol. 3, pp 233-254.
2. Doyle, H. A.; Mamula, M. J., Post-translational protein modifications in antigen recognition and autoimmunity. *Trends in Immunology* **2001,** *22* (8), 443-449.
3. Hirano, H., Microsequence analysis of winged bean seed proteins electroblotted from two-dimensional gel. *Journal of Protein Chemistry* **1989,** *8* (1), 115-130.
4. Abraham, P. E.; Garcia, B. J.; Gunter, L. E.; Jawdy, S. S.; Engle, N.; Yang, X.; Jacobson, D. A.; Hettich, R. L.; Tuskan, G. A.; Tschaplinski, T. J., Quantitative proteome profile of water deficit stress responses in eastern cottonwood (Populus deltoides) leaves. *PLOS ONE* **2018,** *13* (2), e0190019.
5. Rosa, R. L.; Berger, M.; Santi, L.; Driemeier, D.; Barros Terraciano, P.; Campos, A. R.; Guimarães, J. A.; Vainstein, M. H.; Yates, J. R.; Beys-da-Silva, W. O., Proteomics of Rat Lungs Infected by Cryptococcus gattii Reveals a Potential Warburg-like Effect. *Journal of Proteome Research* **2019,** *18* (11), 3885-3895.
6. Tao, D.; Qiao, X.; Sun, L.; Hou, C.; Gao, L.; Zhang, L.; Shan, Y.; Liang, Z.; Zhang, Y., Development of a Highly Efficient 2-D System with a Serially Coupled Long Column and Its Application in Identification of Rat Brain Integral Membrane Proteins with Ionic Liquids-Assisted Solubilization and Digestion. *Journal of Proteome Research* **2011,** *10* (2), 732-738.
7. Szymanski, J.; Levin, Y.; Savidor, A.; Breitel, D.; Chappell-Maor, L.; Heinig, U.; Töpfer, N.; Aharoni, A., Label-free deep shotgun proteomics reveals protein dynamics during tomato fruit tissues development. *The Plant Journal* **2017,** *90* (2), 396-417.
8. Nagaraj, N.; Alexander Kulak, N.; Cox, J.; Neuhauser, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M., System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Molecular & Cellular Proteomics* **2012,** *11* (3), M111.013722.
9. Nesvizhskii, A. I., Proteogenomics: concepts, applications and computational strategies. *Nature Methods* **2014,** *11* (11), 1114-1125.
10. McDonald, W. H.; Yates, J. R., Shotgun Proteomics and Biomarker Discovery. *Disease Markers* **2002,** *18*, 99-105.
11. Cutler, P.; Voshol, H., Proteomics in pharmaceutical research and development. *PROTEOMICS – Clinical Applications* **2015,** *9* (7-8), 643-650.
12. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **2007,** *389* (4), 1017-1031.
13. Winter, G.; Krömer, J. O., Fluxomics – connecting 'omics analysis and phenotypes. *Environmental Microbiology* **2013,** *15* (7), 1901-1916.
14. Krömer, J. O.; Sorgenfrei, O.; Klopprogge, K.; Heinzle, E.; Wittmann, C., In-Depth Profiling of Lysine-Producing Corynebacterium by Combined Analysis of the Transcriptome, Metabolome, and Fluxome. *Journal of Bacteriology* **2004,** *186* (6), 1769.
15. Fiehn, O., Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology* **2002,** *48* (1), 155-171.
16. Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.;

Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L., HMDB: the Human Metabolome Database. *Nucleic Acids Res* **2007,** *35* (Database issue), D521-6.
17. Kind, T.; Fiehn, O., Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **2006,** *7* (1), 234.
18. Lei, Z.; Huhman, D. V.; Sumner, L. W., Mass Spectrometry Strategies in Metabolomics. *Journal of Biological Chemistry* **2011,** *286* (29), 25435-25442.
19. Ramautar, R.; Somsen, G. W.; de Jong, G. J., CE-MS in metabolomics. *Electrophoresis* **2009,** *30* (1), 276-291.
20. Vallejo, M.; Angulo, S.; García-Martínez, D.; García, A.; Barbas, C., New perspective of diabetes response to an antioxidant treatment through metabolic fingerprinting of urine by capillary electrophoresis. *Journal of Chromatography A* **2008,** *1187* (1), 267-274.
21. Monton, M. R. N.; Soga, T., Metabolome analysis by capillary electrophoresis–mass spectrometry. *Journal of Chromatography A* **2007,** *1168* (1), 237-246.
22. Hannun, Y. A.; Obeid, L. M., Principles of bioactive lipid signalling: lessons from sphingolipids. *Nature Reviews Molecular Cell Biology* **2008,** *9* (2), 139-150.
23. Wymann, M. P.; Schneiter, R., Lipid signalling in disease. *Nature Reviews Molecular Cell Biology* **2008,** *9* (2), 162-176.
24. Irvine, R. F., Nuclear lipid signalling. *Nature Reviews Molecular Cell Biology* **2003,** *4* (5), 349-361.
25. Hu, C.; van Dommelen, J.; van der Heijden, R.; Spijksma, G.; Reijmers, T. H.; Wang, M.; Slee, E.; Lu, X.; Xu, G.; van der Greef, J.; Hankemeier, T., RPLC-Ion-Trap-FTMS Method for Lipid Profiling of Plasma: Method Validation and Application to p53 Mutant Mouse Model. *Journal of Proteome Research* **2008,** *7* (11), 4982-4991.
26. Hines, K. M.; Herron, J.; Xu, L., Assessment of altered lipid homeostasis by HILIC-ion mobility-mass spectrometry-based lipidomics. *Journal of Lipid Research* **2017,** *58* (4), 809-819.
27. Han, X.; Gross, R. W., Electrospray ionization mass spectroscopic analysis of human erythrocyte plasma membrane phospholipids. *Proceedings of the National Academy of Sciences* **1994,** *91* (22), 10635.
28. Han, X.; Gross, R. W., Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *Journal of Lipid Research* **2003,** *44* (6), 1071-1079.
29. Cai, S.-S.; Short, L. C.; Syage, J. A.; Potvin, M.; Curtis, J. M., Liquid chromatography–atmospheric pressure photoionization-mass spectrometry analysis of triacylglycerol lipids—Effects of mobile phases on sensitivity. *Journal of Chromatography A* **2007,** *1173* (1), 88-97.
30. Sommer, U.; Herscovitz, H.; Welty, F. K.; Costello, C. E., LC-MS-based method for the qualitative and quantitative analysis of complex lipid mixtures. *Journal of Lipid Research* **2006,** *47* (4), 804-814.
31. Hutchins, P. M.; Barkley, R. M.; Murphy, R. C., Separation of cellular nonpolar neutral lipids by normal-phase chromatography and analysis by electrospray ionization mass spectrometry. *Journal of Lipid Research* **2008,** *49* (4), 804-813.
32. Kind, T.; Fiehn, O., Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **2007,** *8* (1), 105.
33. Kind, T.; Wohlgemuth, G.; Lee, D. Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O., FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole

and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Analytical Chemistry* **2009,** *81* (24), 10038-10048.
34. Bajad, S. U.; Lu, W.; Kimball, E. H.; Yuan, J.; Peterson, C.; Rabinowitz, J. D., Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *Journal of Chromatography A* **2006,** *1125* (1), 76-88.
35. Lu, W.; Clasquin, M. F.; Melamud, E.; Amador-Noguez, D.; Caudy, A. A.; Rabinowitz, J. D., Metabolomic Analysis via Reversed-Phase Ion-Pairing Liquid Chromatography Coupled to a Stand Alone Orbitrap Mass Spectrometer. *Analytical Chemistry* **2010,** *82* (8), 3212-3221.
36. Ansó, E.; Weinberg, S. E.; Diebold, L. P.; Thompson, B. J.; Malinge, S.; Schumacker, Paul T.; Liu, X.; Zhang, Y.; Shao, Z.; Steadman, M.; Marsh, K. M.; Xu, J.; Crispino, John D.; Chandel, N. S., The mitochondrial respiratory chain is essential for haematopoietic stem cell function. *Nature Cell Biology* **2017,** *19* (6), 614-625.
37. Zhang, Y.; Kurupati, R.; Liu, L.; Zhou, X. Y.; Zhang, G.; Hudaihed, A.; Filisio, F.; Giles-Davis, W.; Xu, X.; Karakousis, G. C.; Schuchter, L. M.; Xu, W.; Amaravadi, R.; Xiao, M.; Sadek, N.; Krepler, C.; Herlyn, M.; Freeman, G. J.; Rabinowitz, J. D.; Ertl, H. C. J., Enhancing CD8+ T Cell Fatty Acid Catabolism within a Metabolically Challenging Tumor Microenvironment Increases the Efficacy of Melanoma Immunotherapy. *Cancer Cell* **2017,** *32* (3), 377-391.e9.
38. Clemmons, B. A.; Martino, C.; Powers, J. B.; Campagna, S. R.; Voy, B. H.; Donohoe, D. R.; Gaffney, J.; Embree, M. M.; Myer, P. R., Rumen Bacteria and Serum Metabolites Predictive of Feed Efficiency Phenotypes in Beef Cattle. *Scientific Reports* **2019,** *9* (1), 19265.
39. Virtue, A. T.; McCright, S. J.; Wright, J. M.; Jimenez, M. T.; Mowel, W. K.; Kotzin, J. J.; Joannas, L.; Basavappa, M. G.; Spencer, S. P.; Clark, M. L.; Eisennagel, S. H.; Williams, A.; Levy, M.; Manne, S.; Henrickson, S. E.; Wherry, E. J.; Thaiss, C. A.; Elinav, E.; Henao-Mejia, J., The gut microbiota regulates white adipose tissue inflammation and obesity via a family of microRNAs. *Science Translational Medicine* **2019,** *11* (496), eaav1892.
40. Krycer, J. R.; Quek, L.-E.; Francis, D.; Fazakerley, D. J.; Elkington, S. D.; Diaz-Vegas, A.; Cooke, K. C.; Weiss, F. C.; Duan, X.; Kurdyukov, S.; Zhou, P.-X.; Tambar, U. K.; Hirayama, A.; Ikeda, S.; Kamei, Y.; Soga, T.; Cooney, G. J.; James, D. E., Lactate production is a prioritized feature of adipocyte metabolism. *Journal of Biological Chemistry* **2020,** *295* (1), 83-98.
41. Mushtaq, M. Y.; Choi, Y. H.; Verpoorte, R.; Wilson, E. G., Extraction for Metabolomics: Access to The Metabolome. *Phytochemical Analysis* **2014,** *25* (4), 291-306.
42. Koning, W. d.; Dam, K. v., A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Analytical Biochemistry* **1992,** *204* (1), 118-123.
43. Hajjaj, H.; Blanc, P. J.; Goma, G.; François, J., Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi. *FEMS Microbiology Letters* **1998,** *164* (1), 195-200.
44. Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G., Solvent-Dependent Metabolite Distribution, Clustering, and Protein Extraction for Serum Profiling with Mass Spectrometry. *Analytical Chemistry* **2006,** *78* (3), 743-752.

45. Gil, A.; Siegel, D.; Permentier, H.; Reijngoud, D.-J.; Dekker, F.; Bischoff, R., Stability of energy metabolites—An often overlooked issue in metabolomics studies: A review. *Electrophoresis* **2015,** *36* (18), 2156-2169.
46. Dietmair, S.; Timmins, N. E.; Gray, P. P.; Nielsen, L. K.; Krömer, J. O., Towards quantitative metabolomics of mammalian cells: Development of a metabolite extraction protocol. *Analytical Biochemistry* **2010,** *404* (2), 155-164.
47. Winder, C. L.; Dunn, W. B.; Schuler, S.; Broadhurst, D.; Jarvis, R.; Stephens, G. M.; Goodacre, R., Global Metabolic Profiling of Escherichia coli Cultures: an Evaluation of Methods for Quenching and Extraction of Intracellular Metabolites. *Analytical Chemistry* **2008,** *80* (8), 2939-2948.
48. Hancock, R., The intracellular amino acids of Staphylococcus aureus: Release and analysis. *Biochimica et Biophysica Acta* **1958,** *28*, 402-412.
49. Freeland, J. C.; Gale, E. F., The Amio-acid Composition of Certain Bacteria and Yeasts. *Biochemical Journal* **1947,** *41* (1), 135-138.
50. Bent, K. J.; Morton, A. G., Amino Acid Composition of Fungi during Development in Submerged Culture. *Biochemical Journal* **1964,** *92* (260), 260-269.
51. Canelas, A. B.; ten Pierick, A.; Ras, C.; Seifar, R. M.; van Dam, J. C.; van Gulik, W. M.; Heijnen, J. J., Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics. *Analytical Chemistry* **2009,** *81* (17), 7379-7389.
52. Prasad Maharjan, R.; Ferenci, T., Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli. *Analytical Biochemistry* **2003,** *313* (1), 145-154.
53. Rabinowitz, J. D.; Kimball, E., Acidic Acetonitrile for Cellular Metabolome Extraction from Escherichia coli. *Analytical Chemistry* **2007,** *79* (16), 6167-6173.
54. Bragger, J. M.; Dunn, R. V.; Daniel, R. M., Enzyme activity down to −100°C. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **2000,** *1480* (1), 278-282.
55. Folch, J.; Ascoli, I.; Lees, M.; Meath, J.; LeBaron, F., Preparation of Lipide Extracts from Brain Tissue. *Journal of Biological Chemistry* **1951,** (191), 833-841.
56. Folch, J. M.; Stanley, S. G. H., A simple method for the isolation and purification of total lipides from animal issues. *Journal of Biological Chemistry* **1957,** *226* (1), 497-509.
57. Bligh, E. G.; Dyer, W. J., A Rapid Method of Total Lipid Extraction and Purification. *Canadian Journal of Biochemistry and Physiology* **1959,** *37* (1), 911-917.
58. Iverson, S. J.; Lang, S. L. C.; Cooper, M. H., Comparison of the bligh and dyer and folch methods for total lipid determination in a broad range of marine tissue. *Lipids* **2001,** *36* (11), 1283-1287.
59. Stefely, J. A.; Kwiecien, N. W.; Freiberger, E. C.; Richards, A. L.; Jochem, A.; Rush, M. J. P.; Ulbrich, A.; Robinson, K. P.; Hutchins, P. D.; Veling, M. T.; Guo, X.; Kemmerer, Z. A.; Connors, K. J.; Trujillo, E. A.; Sokol, J.; Marx, H.; Westphall, M. S.; Hebert, A. S.; Pagliarini, D. J.; Coon, J. J., Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. *Nature Biotechnology* **2016,** *34* (11), 1191-1197.
60. Coman, C.; Solari, F. A.; Hentschel, A.; Sickmann, A.; Zahedi, R. P.; Ahrends, R., Simultaneous Metabolite, Protein, Lipid Extraction (SIMPLEX): A Combinatorial Multimolecular Omics Approach for Systems Biology. *Molecular & Cellular Proteomics* **2016,** *15* (4), 1453.

61. Nakayasu, E. S.; Nicora, C. D.; Sims, A. C.; Burnum-Johnson, K. E.; Kim, Y.-M.; Kyle, J. E.; Matzke, M. M.; Shukla, A. K.; Chu, R. K.; Schepmoes, A. A.; Jacobs, J. M.; Baric, R. S.; Webb-Robertson, B.-J.; Smith, R. D.; Metz, T. O., MPLEx: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. *mSystems* **2016,** *1* (3), e00043-16.
62. Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D., Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *Journal of Lipid Research* **2008,** *49* (5), 1137-1146.
63. Horning, E. C.; Horning, M. G., Human Metabolic Profiles Obtained by GC and GC/MS. *Journal of Chromatographic Science* **1971,** *9* (3), 129-140.
64. Horning, E. C.; Horning, M. G., Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites. *Clinical Chemistry* **1971,** *17* (8), 802.
65. Mrochek, J. E.; Butts, W. C.; Rainey, W. T.; Burtis, C. A., Separation and Identification of Urinary Constituents by Use of Multiple-Analytical Techniques. *Clinical Chemistry* **1971,** *17* (2), 72.
66. Thompson, J. A.; Markey, S. P., Quantitative metabolic profiling of urinary organic acids by gas chromatography-mass spectrometry. Comparison of isolation methods. *Analytical Chemistry* **1975,** *47* (8), 1313-1321.
67. Fiehn, O., Metabolomics by Gas Chromatography–Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Current Protocols in Molecular Biology* **2016,** *114* (1), 30.4.1-30.4.32.
68. Seo, C.; Hwang, Y.-H.; Lee, H.-S.; Kim, Y.; Shin, T. H.; Lee, G.; Son, Y.-J.; Kim, H.; Yee, S.-T.; Park, A. K.; Paik, M.-J., Metabolomic study for monitoring of biomarkers in mouse plasma with asthma by gas chromatography–mass spectrometry. *Journal of Chromatography B* **2017,** *1063*, 156-162.
69. Higgins Keppler, E. A.; Jenkins, C. L.; Davis, T. J.; Bean, H. D., Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics. *TrAC Trends in Analytical Chemistry* **2018,** *109*, 275-286.
70. Careri, M.; Elviri, L.; Mangia, A., Liquid chromatography–electrospray mass spectrometry of β-carotene and xanthophylls: Validation of the analytical method. *Journal of Chromatography A* **1999,** *854* (1), 233-244.
71. Plumb, R.; Granger, J.; Stumpf, C.; Wilson, I. D.; Evans, J. A.; Lenz, E. M., Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst* **2003,** *128* (7), 819-823.
72. Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N., Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development. *Rapid Communications in Mass Spectrometry* **2002,** *16* (20), 1991-1996.
73. Welti, R.; Li, W.; Li, M.; Sang, Y.; Biesiada, H.; Zhou, H.-E.; Rajashekar, C. B.; Williams, T. D.; Wang, X., Profiling Membrane Lipids in Plant Stress Responses: ROLE OF PHOSPHOLIPASE Dα IN FREEZING-INDUCED LIPID CHANGES IN ARABIDOPSIS. *Journal of Biological Chemistry* **2002,** *277* (35), 31994-32002.

74. Lu, W.; Kimball, E.; Rabinowitz, J. D., A High-Performance Liquid Chromatography-Tandem Mass Spectrometry Method for Quantitation of Nitrogen-Containing Intracellular Metabolites. *Journal of the American Society for Mass Spectrometry* **2006,** *17* (1), 37-50.
75. Lu, W.; Bennett, B. D.; Rabinowitz, J. D., Analytical strategies for LC–MS-based targeted metabolomics. *Journal of Chromatography B* **2008,** *871* (2), 236-242.
76. McCalley, D. V.; Neue, U. D., Estimation of the extent of the water-rich layer associated with the silica surface in hydrophilic interaction chromatography. *Journal of Chromatography A* **2008,** *1192* (2), 225-229.
77. McCalley, D. V., Study of the selectivity, retention mechanisms and performance of alternative silica-based stationary phases for separation of ionised solutes in hydrophilic interaction chromatography. *Journal of Chromatography A* **2010,** *1217* (20), 3408-3417.
78. Luo, B.; Groenke, K.; Takors, R.; Wandrey, C.; Oldiges, M., Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography–mass spectrometry. *Journal of Chromatography A* **2007,** *1147* (2), 153-164.
79. Tague, E. D.; Bourdon, A. K.; MacDonald, A.; Lookadoo, M. S.; Kim, E. D.; White, W. M.; Terry, P. D.; Campagna, S. R.; Voy, B. H.; Whelan, J., Metabolomics Approach in the Study of the Well-Defined Polyherbal Preparation Zyflamend. *Journal of Medicinal Food* **2017,** *21* (3), 306-316.
80. Wells, A.; Barrington, W. T.; Dearth, S.; May, A.; Threadgill, D. W.; Campagna, S. R.; Voy, B. H., Tissue Level Diet and Sex-by-Diet Interactions Reveal Unique Metabolite and Clustering Profiles Using Untargeted Liquid Chromatography–Mass Spectrometry on Adipose, Skeletal Muscle, and Liver Tissue in C57BL6/J Mice. *Journal of Proteome Research* **2018,** *17* (3), 1077-1090.
81. Powers, J. B.; Campagna, S. R., Design and Evaluation of a Gas Chromatograph-Atmospheric Pressure Chemical Ionization Interface for an Exactive Orbitrap Mass Spectrometer. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (11), 2369-2379.
82. Li, M.; Yang, L.; Bai, Y.; Liu, H., Analytical Methods in Lipidomics and Their Applications. *Analytical Chemistry* **2014,** *86* (1), 161-175.
83. Han, X.; Gross, R. W., Quantitative Analysis and Molecular Species Fingerprinting of Triacylglyceride Molecular Species Directly from Lipid Extracts of Biological Samples by Electrospray Ionization Tandem Mass Spectrometry. *Analytical Biochemistry* **2001,** *295* (1), 88-100.
84. HAN, X.; ABENDSCHEIN, D. R.; KELLEY, J. G.; GROSS, R. W., Diabetes-induced changes in specific lipid molecular species in rat myocardium. *Biochemical Journal* **2000,** *352* (1), 79-89.
85. Ståhlman, M.; Ejsing, C. S.; Tarasov, K.; Perman, J.; Borén, J.; Ekroos, K., High-throughput shotgun lipidomics by quadrupole time-of-flight mass spectrometry. *Journal of Chromatography B* **2009,** *877* (26), 2664-2672.
86. Han, X.; Yang, K.; Gross, R. W., Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrometry Reviews* **2012,** *31* (1), 134-178.
87. Ejsing, C. S.; Sampaio, J. L.; Surendranath, V.; Duchoslav, E.; Ekroos, K.; Klemm, R. W.; Simons, K.; Shevchenko, A., Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proceedings of the National Academy of Sciences* **2009,** *106* (7), 2136.

88. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry* **2006,** *78* (3), 779-787.
89. Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S., CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Analytical Chemistry* **2012,** *84* (1), 283-289.
90. Daly, R.; Rogers, S.; Wandy, J.; Jankevics, A.; Burgess, K. E. V.; Breitling, R., MetAssign: probabilistic annotation of metabolites from LC–MS data using a Bayesian clustering approach. *Bioinformatics* **2014,** *30* (19), 2764-2771.
91. Alonso, A.; Julià, A.; Beltran, A.; Vinaixa, M.; Díaz, M.; Ibañez, L.; Correig, X.; Marsal, S., AStream: an R package for annotating LC/MS metabolomic data. *Bioinformatics* **2011,** *27* (9), 1339-1340.
92. Melamud, E.; Vastag, L.; Rabinowitz, J. D., Metabolomic Analysis and Visualization Engine for LC−MS Data. *Analytical Chemistry* **2010,** *82* (23), 9818-9826.
93. Karp, P. D.; Riley, M.; Paley, S. M.; Pellegrini-Toole, A., The MetaCyc Database. *Nucleic Acids Research* **2002,** *30* (1), 59-61.
94. Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H., Jr.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S., LMSD: LIPID MAPS structure database. *Nucleic Acids Research* **2006,** *35* (suppl_1), D527-D532.
95. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry* **2010,** *45* (7), 703-714.
96. Kanehisa, M.; Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **2000,** *28* (1), 27-30.
97. Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A., HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research* **2017,** *46* (D1), D608-D617.
98. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E., PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research* **2018,** *47* (D1), D1102-D1109.
99. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J., MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* **2018,** *46* (W1), W486-W494.
100. Hogan, S. R.; Phan, J. H.; Alvarado-Velez, M.; Wang, M. D.; Bellamkonda, R. V.; Fernández, F. M.; LaPlaca, M. C., Discovery of Lipidome Alterations Following Traumatic Brain Injury via High-Resolution Metabolomics. *Journal of Proteome Research* **2018,** *17* (6), 2131-2143.
101. Han, X.; Rozen, S.; Boyle, S. H.; Hellegers, C.; Cheng, H.; Burke, J. R.; Welsh-Bohmer, K. A.; Doraiswamy, P. M.; Kaddurah-Daouk, R., Metabolomics in Early Alzheimer's Disease: Identification of Altered Plasma Sphingolipidome Using Shotgun Lipidomics. *PLOS ONE* **2011,** *6* (7), e21643.

102. Kim, K.; Aronov, P.; Zakharkin, S. O.; Anderson, D.; Perroud, B.; Thompson, I. M.; Weiss, R. H., Urine Metabolomics Analysis for Kidney Cancer Detection and Biomarker Discovery. *Molecular & Cellular Proteomics* **2009,** *8* (3), 558.
103. Brockman, S. A.; Roden, E. V.; Hegeman, A. D., Van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. *Metabolomics* **2018,** *14* (4), 48.
104. Vital Signs: Staph infections can kill. https://www.cdc.gov/vitalsigns/staph/index.html (accessed March 25).
105. Jean-Pierre, H.; Darbas, H.; Jean-Roussenq, A.; Boyer, G., Pathogenicity in two cases of Staphylococcus schleiferi, a recently described species. *J Clin Microbiol* **1989,** *27* (9), 2110-1.
106. Latorre, M.; Rojo, P. M.; Unzaga, M. J.; Cisterna, R., Staphylococcus schleiferi: a new opportunistic pathogen. *Clin Infect Dis* **1993,** *16* (4), 589-90.
107. Yarbrough, M. L.; Hamad, Y.; Burnham, C. A.; George, I. A., Closing the Brief Case: Bacteremia and Vertebral Osteomyelitis Due to Staphylococcus schleiferi. *J Clin Microbiol* **2017,** *55* (11), 3309-3310.
108. Yarbrough, M. L.; Hamad, Y.; Burnham, C. A.; George, I. A., The Brief Case: Bacteremia and Vertebral Osteomyelitis Due to Staphylococcus schleiferi. *J Clin Microbiol* **2017,** *55* (11), 3157-3161.
109. Jin, D.; Zhang, S.; Li, M., Staphylococcus schleiferi Meningitis in an Infant. *Pediatr Infect Dis J* **2017,** *36* (2), 243-244.
110. Kumar, D.; Cawley, J. J.; Irizarry-Alvarado, J. M.; Alvarez, A.; Alvarez, S., Case of Staphylococcus schleiferi subspecies coagulans endocarditis and metastatic infection in an immune compromised host. *Transpl Infect Dis* **2007,** *9* (4), 336-8.
111. Frank, L. A.; Kania, S. A.; Hnilica, K. A.; Wilkes, R. P.; Bemis, D. A., Isolation of Staphylococcus schleiferi from dogs with pyoderma. *J Am Vet Med Assoc* **2003,** *222* (4), 451-4.
112. Tzamalis, A.; Chalvatzis, N.; Anastasopoulos, E.; Tzetzi, D.; Dimitrakos, S., Acute postoperative Staphylococcus schleiferi endophthalmitis following uncomplicated cataract surgery: first report in the literature. *Eur J Ophthalmol* **2013,** *23* (3), 427-30.
113. Savini, V.; Barbarini, D.; Polakowska, K.; Gherardi, G.; Białecka, A.; Kasprowicz, A.; Polilli, E.; Marrollo, R.; Di Bonaventura, G.; Fazii, P.; Antonio, D.; Międzobrodzki, J.; Carretto, E., Methicillin-Resistant Staphylococcus pseudintermedius Infection in a Bone Marrow Transplant Recipient. *Journal of Clinical Microbiology* **2013,** *51* (5), 1636.
114. Van Hoovels, L.; Vankeerberghen, A.; Boel, A.; Van Vaerenbergh, K.; De Beenhouwer, H., First Case of Staphylococcus pseudintermedius Infection in a Human. *Journal of Clinical Microbiology* **2006,** *44* (12), 4609.
115. Calvo, J.; Hernández, J. L.; Fariñas, M. C.; García-Palomo, D.; Agüero, J., Osteomyelitis Caused by Staphylococcus schleiferi and Evidence of Misidentification of This Staphylococcus Species by an Automated Bacterial Identification System. *Journal of Clinical Microbiology* **2000,** *38* (10), 3887.
116. Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R., Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences* **2000,** *97* (24), 13172.
117. Voynova, N. E.; Rios, S. E.; Miziorko, H. M., Staphylococcus aureus mevalonate kinase: isolation and characterization of an enzyme of the isoprenoid biosynthetic pathway. *J Bacteriol* **2004,** *186* (1), 61-7.

118. Wilding, E. I.; Brown, J. R.; Bryant, A. P.; Chalker, A. F.; Holmes, D. J.; Ingraham, K. A.; Iordanescu, S.; So, C. Y.; Rosenberg, M.; Gwynn, M. N., Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol* **2000,** *182* (15), 4319-27.
119. Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R., Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A* **2000,** *97* (24), 13172-7.
120. Voynova, N. E.; Rios, S. E.; Miziorko, H. M., Staphylococcus aureus Mevalonate Kinase: Isolation and Characterization of an Enzyme of the Isoprenoid Biosynthetic Pathway. *Journal of Bacteriology* **2004,** *186* (1), 61.
121. Misic, A. M.; Cain, C. L.; Morris, D. O.; Rankin, S. C.; Beiting, D. P., Divergent Isoprenoid Biosynthesis Pathways in Staphylococcus; Species Constitute a Drug Target for Treating Infections in Companion Animals. *mSphere* **2016,** *1* (5), e00258-16.
122. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012,** *28* (12), 1647-9.
123. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G., QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **2013,** *29* (8), 1072-5.
124. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S. J.; Marra, M. A., Circos: an information aesthetic for comparative genomics. *Genome Res* **2009,** *19* (9), 1639-45.
125. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L., BLAST+: architecture and applications. *BMC Bioinformatics* **2009,** *10*, 421.
126. Chen, H.; Boutros, P. C., VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **2011,** *12*, 35.
127. Zhou, Y.; Liang, Y.; Lynch, K. H.; Dennis, J. J.; Wishart, D. S., PHAST: a fast phage search tool. *Nucleic Acids Res* **2011,** *39* (Web Server issue), W347-52.
128. Bland, C.; Ramsey, T. L.; Sabree, F.; Lowe, M.; Brown, K.; Kyrpides, N. C.; Hugenholtz, P., CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **2007,** *8*, 209.
129. Aziz, R. K.; Bartels, D.; Best, A. A.; DeJongh, M.; Disz, T.; Edwards, R. A.; Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; Meyer, F.; Olsen, G. J.; Olson, R.; Osterman, A. L.; Overbeek, R. A.; McNeil, L. K.; Paarmann, D.; Paczian, T.; Parrello, B.; Pusch, G. D.; Reich, C.; Stevens, R.; Vassieva, O.; Vonstein, V.; Wilke, A.; Zagnitko, O., The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **2008,** *9*, 75.
130. Dearth, S. P.; Castro, H. F.; Venice, F.; Tague, E. D.; Novero, M.; Bonfante, P.; Campagna, S. R., Metabolome changes are induced in the arbuscular mycorrhizal fungus Gigaspora margarita by germination and by its bacterial endosymbiont. *Mycorrhiza* **2018,** *28* (5), 421-433.
131. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.;

Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **2012,** *30*, 918.
132. Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Römpp, A.; Neumann, S.; Pizarro, A. D.; Montecchi-Palazzi, L.; Tasman, N.; Coleman, M.; Reisinger, F.; Souda, P.; Hermjakob, H.; Binz, P.-A.; Deutsch, E. W., mzML—a Community Standard for Mass Spectrometry Data. *Molecular & Cellular Proteomics* **2011,** *10* (1), R110.000133.
133. Misic, A. M.; Cain, C. L.; Morris, D. O.; Rankin, S. C.; Beiting, D. P., Divergent Isoprenoid Biosynthesis Pathways in Staphylococcus Species Constitute a Drug Target for Treating Infections in Companion Animals. *mSphere* **2016,** *1* (5), e00258-16.
134. Misic, A. M.; Cain, C. L.; Morris, D. O.; Rankin, S. C.; Beiting, D. P., Divergent Isoprenoid Biosynthesis Pathways in Staphylococcus Species Constitute a Drug Target for Treating Infections in Companion Animals. *mSphere* **2016,** *1* (5).
135. Bordbar, A.; Mo, M. L.; Nakayasu, E. S.; Schrimpe-Rutledge, A. C.; Kim, Y.-M.; Metz, T. O.; Jones, M. B.; Frank, B. C.; Smith, R. D.; Peterson, S. N.; Hyduke, D. R.; Adkins, J. N.; Palsson, B. O., Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Molecular Systems Biology* **2012,** *8* (1), 558.
136. Hultman, J.; Waldrop, M. P.; Mackelprang, R.; David, M. M.; McFarland, J.; Blazewicz, S. J.; Harden, J.; Turetsky, M. R.; McGuire, A. D.; Shah, M. B.; VerBerkmoes, N. C.; Lee, L. H.; Mavrommatis, K.; Jansson, J. K., Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **2015,** *521* (7551), 208-212.
137. Faijes, M.; Mars, A. E.; Smid, E. J., Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum. *Microbial Cell Factories* **2007,** *6* (1), 27.
138. Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M., Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis* **2010,** *31* (21), 3573-3579.
139. Weckwerth, W.; Wenzel, K.; Fiehn, O., Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *PROTEOMICS* **2004,** *4* (1), 78-83.
140. Belle, J. E. L.; Harris, N. G.; Williams, S. R.; Bhakoo, K. K., A comparison of cell and tissue extraction techniques using high-resolution 1H-NMR spectroscopy. *NMR in Biomedicine* **2002,** *15* (1), 37-44.
141. Roume, H.; El Muller, E.; Cordes, T.; Renaut, J.; Hiller, K.; Wilmes, P., A biomolecular isolation framework for eco-systems biology. *The ISME Journal* **2013,** *7* (1), 110-121.
142. Nelson, K. E.; Weinel, C.; Paulsen, I. T.; Dodson, R. J.; Hilbert, H.; Martins dos Santos, V. A. P.; Fouts, D. E.; Gill, S. R.; Pop, M.; Holmes, M.; Brinkac, L.; Beanan, M.; DeBoy, R. T.; Daugherty, S.; Kolonay, J.; Madupu, R.; Nelson, W.; White, O.; Peterson, J.; Khouri, H.; Hance, I.; Lee, P. C.; Holtzapple, E.; Scanlan, D.; Tran, K.; Moazzez, A.; Utterback, T.; Rizzo, M.; Lee, K.; Kosack, D.; Moestl, D.; Wedler, H.; Lauber, J.; Stjepandic, D.; Hoheisel, J.; Straetz, M.; Heim, S.; Kiewitz, C.; Eisen, J.; Timmis, K. N.; Düsterhöft, A.; Tümmler, B.; Fraser, C. M., Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environmental Microbiology* **2002,** *4* (12), 799-808.
143. Ramos-González, M. I.; Campos, M. J.; Ramos, J. L., Analysis of Pseudomonas putida KT2440 gene expression in the maize rhizosphere: in vivo [corrected] expression technology

capture and identification of root-activated promoters. *Journal of bacteriology* **2005,** *187* (12), 4033-4041.
144. Nwachukwu, S. U., Bioremediation of Sterile Agricultural Soils Polluted with Crude Petroleum by Application of the Soil Bacterium, Pseudomonas putida, with Inorganic Nutrient Supplementations. *Current Microbiology* **2001,** *42* (4), 231-236.
145. Zuo, Z.; Gong, T.; Che, Y.; Liu, R.; Xu, P.; Jiang, H.; Qiao, C.; Song, C.; Yang, C., Engineering Pseudomonas putida KT2440 for simultaneous degradation of organophosphates and pyrethroids and its application in bioremediation of soil. *Biodegradation* **2015,** *26* (3), 223-233.
146. Chapatwala, K. D.; Babu, G. R. V.; Armstead, E. R.; White, E. M.; Wolfram, J. H., A kinetic study on the bioremediation of sodium cyanide and acetonitrile by free and immobilized cells ofPseudomonas putida. *Applied Biochemistry and Biotechnology* **1995,** *51* (1), 717-726.
147. Samin, G.; Pavlova, M.; Arif, M. I.; Postema, C. P.; Damborsky, J.; Janssen, D. B., A Pseudomonas putida Strain Genetically Engineered for 1,2,3-Trichloropropane Bioremediation. *Applied and Environmental Microbiology* **2014,** *80* (17), 5467.
148. Khan, H. A.; Ahmad, A.; Mehboob, R., Nosocomial infections and their control strategies. *Asian Pacific Journal of Tropical Biomedicine* **2015,** *5* (7), 509-514.
149. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*, 395.
150. Kondakova, T.; D'Heygère, F.; Feuilloley, M. J.; Orange, N.; Heipieper, H. J.; Duclairoir Poc, C., Glycerophospholipid synthesis and functions in Pseudomonas. *Chemistry and Physics of Lipids* **2015,** *190*, 27-42.
151. Rühl, J.; Hein, E.-M.; Hayen, H.; Schmid, A.; Blank, L. M., The glycerophospholipid inventory of Pseudomonas putida is conserved between strains and enables growth condition-related alterations. *Microbial Biotechnology* **2012,** *5* (1), 45-58.
152. Löfgren, L.; Forsberg, G.-B.; Ståhlman, M., The BUME method: a new rapid and simple chloroform-free method for total lipid extraction of animal tissue. *Scientific Reports* **2016,** *6* (1), 27688.
153. Bos, J. L.; Rehmann, H.; Wittinghofer, A., GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell* **2007,** *129* (5), 865-877.
154. Cordell, R. L.; Hill, S. J.; Ortori, C. A.; Barrett, D. A., Quantitative profiling of nucleotides and related phosphate-containing metabolites in cultured mammalian cells by liquid chromatography tandem electrospray mass spectrometry. *Journal of Chromatography B* **2008,** *871* (1), 115-124.

Appendix

Table A-1. Staphylococcus Metabolites Normalized 196, 192, 06-3228. This table lists the normalized intensities associated with metabolite annotations for strains 196, 192, and 06-3228. 196_1 had an optical density of 0.402, 196_2 had an optical density of 0.415, 196_3 had an optical density of 0.470, 192_1 had an optical density of 0.434, 192_2 had an optical density of 0.447, 192_3 had an optical density of 0.445, 06-3228_1 had an optical density of 0.338, 06-3228_2 had an optical density of 0.349, and 06-3228_3 had an optical density of 0.337.

COMPOUND	196_1	196_2	196_3	192_1	192_2	192_3	06-3228_1	06-3228_2	06-3228_3
Dimethylglycine	4.6E+07	4.2E+07	7.0E+07	3.1E+07	2.1E+07	4.4E+07	4.9E+07	6.7E+07	5.8E+07
hydroxybutyrate	1.9E+08	2.3E+08	1.9E+08	1.8E+08	2.6E+08	3.4E+08	2.8E+08	2.4E+08	2.7E+08
Histamine	8.7E+03	2.8E+04	2.1E+04	8.6E+03	9.9E+03	1.2E+04	4.5E+04	3.2E+04	4.8E+04
Proline	1.1E+08	1.3E+08	1.1E+08	6.2E+08	6.3E+08	4.5E+08	8.3E+07	6.7E+07	7.5E+07
Fumarate	7.0E+06	6.5E+07	1.3E+07	4.2E+07	3.5E+07	3.7E+07	1.6E+07	1.6E+07	1.6E+07
2-Oxoisovalerate	1.1E+06	1.1E+06	1.2E+06	2.4E+06	4.0E+06	5.3E+06	6.7E+06	9.6E+06	4.9E+06
Valine/betaine	5.3E+08	8.9E+08	6.1E+08	2.2E+08	3.1E+08	6.5E+08	5.3E+08	5.0E+08	4.6E+08
Succinate/Methylmalonate	1.4E+09	2.1E+09	1.5E+09	9.5E+08	1.4E+09	2.3E+09	1.6E+09	1.5E+09	1.7E+09
3-Hydroxyisovalerate	1.7E+08	1.5E+08	1.8E+08	3.4E+08	5.1E+08	6.7E+08	4.7E+08	4.4E+08	4.3E+08
Homoserine/Threonine	5.3E+07	5.9E+07	5.1E+07	1.5E+08	2.1E+08	2.9E+08	2.0E+07	1.5E+07	2.1E+07
3-Methylthiopropionate	3.6E+05	3.1E+05	2.1E+05	1.6E+06	1.9E+06	2.7E+06	2.7E+06	2.5E+06	2.7E+06
Cysteine	1.1E+06	2.0E+06	1.3E+06	7.1E+05	1.3E+06	2.3E+06	6.5E+05	5.8E+05	7.7E+05
Citraconate	1.9E+07	3.1E+07	1.8E+07	1.7E+07	2.5E+07	4.7E+07	2.3E+07	2.2E+07	1.7E+07
N-Acetylputrescine	2.7E+05	6.2E+05	3.5E+05	4.9E+06	5.0E+06	5.2E+06	2.4E+05	2.0E+05	2.2E+05
Hydroxyproline	7.2E+07	8.5E+07	9.0E+07	1.2E+08	1.2E+08	1.1E+08	1.9E+07	1.7E+07	1.6E+07
Leucine/Isoleucine	3.1E+09	5.2E+09	3.5E+09	2.4E+09	3.1E+09	6.0E+09	2.7E+09	2.7E+09	2.3E+09
methyl succinic acid	9.9E+07	1.3E+08	1.2E+08	6.2E+07	7.9E+07	1.2E+08	6.9E+07	6.8E+07	7.6E+07
Asparagine	7.4E+08	8.4E+08	7.4E+08	2.8E+08	3.5E+08	5.0E+08	9.8E+08	9.4E+08	8.2E+08
Hydroxyisocaproic acid	3.5E+08	3.6E+08	3.7E+08	1.5E+09	2.7E+09	3.4E+09	2.4E+09	2.4E+09	2.3E+09
Ornithine	2.7E+08	2.6E+08	2.6E+08	4.1E+07	4.7E+07	6.1E+07	5.2E+07	5.1E+07	5.2E+07
Aspartate	2.4E+09	3.0E+09	2.8E+09	2.7E+09	3.0E+09	4.1E+09	2.0E+09	2.0E+09	1.6E+09
Homocysteine	7.3E+05	1.0E+06	6.3E+05	5.0E+05	5.2E+05	5.7E+05	2.7E+05	2.7E+05	2.7E+05
Anthranilate	9.9E+06	8.7E+06	1.2E+07	1.1E+07	1.0E+07	9.1E+06	7.2E+06	7.0E+06	7.5E+06
Hypoxanthine	9.5E+07	2.3E+08	1.1E+08	9.1E+06	1.5E+07	5.0E+07	5.9E+07	5.5E+07	5.7E+07
Salicylate	5.3E+07	4.5E+07	6.5E+07	4.3E+07	3.6E+07	2.8E+07	2.8E+07	2.9E+07	2.6E+07

Table A-1. Continued.

COMPOUND	196_1	196_2	196_3	192_1	192_2	192_3	06-3228_1	06-3228_2	06-3228_3
Hydroxybenzoate	2.7E+07	2.8E+07	3.0E+07	3.1E+07	3.0E+07	2.8E+07	1.4E+07	1.5E+07	1.4E+07
alpha-Ketoglutarate	1.6E+07	1.9E+07	2.5E+07	3.2E+07	7.3E+07	1.1E+08	1.5E+07	1.0E+07	1.3E+07
Glutamine	2.7E+08	2.7E+08	3.1E+08	3.8E+08	3.8E+08	3.4E+08	1.4E+08	1.3E+08	7.9E+07
Lysine	2.5E+07	3.5E+07	2.6E+07	2.9E+07	3.6E+07	4.8E+07	4.3E+07	4.4E+07	3.1E+07
O-Acetyl-L-serine	1.4E+07	1.4E+07	1.2E+07	3.1E+06	4.3E+06	2.0E+06	6.9E+06	6.2E+06	4.1E+06
Glutamate	2.6E+09	3.1E+09	2.8E+09	2.0E+09	2.7E+09	3.2E+09	1.5E+09	1.5E+09	1.1E+09
Methionine	1.7E+08	3.5E+08	1.8E+08	1.1E+08	1.7E+08	4.3E+08	2.3E+08	2.1E+08	1.9E+08
Guanine	1.8E+07	3.9E+07	2.1E+07	1.4E+07	2.3E+07	6.4E+07	2.1E+07	1.9E+07	1.8E+07
Vanillin	9.1E+07	8.5E+07	8.9E+07	8.8E+07	7.9E+07	5.8E+07	8.5E+07	9.1E+07	1.0E+08
Xylitol	1.3E+06	6.2E+06	2.3E+06	2.3E+06	3.7E+06	6.6E+06	6.6E+06	6.6E+06	9.1E+06
Orotate	4.4E+07	8.6E+07	7.3E+07	3.0E+07	4.0E+07	1.1E+08	1.7E+07	1.4E+07	2.2E+07
Dihydroorotate	1.9E+08	1.8E+08	1.6E+08	5.3E+06	8.2E+06	2.5E+07	1.1E+07	1.0E+07	1.1E+07
pimelic acid	1.8E+07	1.9E+07	2.1E+07	1.4E+07	1.5E+07	1.5E+07	1.7E+07	1.8E+07	1.8E+07
Indole-3-carboxylate	1.3E+06	1.2E+06	1.4E+06	1.5E+06	1.3E+06	1.3E+06	1.1E+06	9.7E+05	1.1E+06
Phenylpyruvate	5.4E+07	4.1E+07	4.7E+07	4.9E+07	5.7E+07	4.0E+07	1.3E+07	8.9E+06	1.2E+07
Methionine sulfoxide	8.2E+06	2.0E+07	1.1E+07	7.3E+06	1.0E+07	2.5E+07	1.4E+07	1.4E+07	1.3E+07
Phenyllactic acid	1.1E+08	1.2E+08	1.3E+08	3.5E+08	4.5E+08	5.5E+08	2.1E+08	2.0E+08	2.0E+08
Cysteate	1.2E+08	1.3E+08	1.5E+08	5.0E+07	5.3E+07	6.8E+07	3.1E+07	2.5E+07	3.3E+07
Sulfolactate	4.9E+07	1.4E+08	6.9E+07	2.6E+07	4.4E+07	1.4E+08	8.8E+07	8.4E+07	8.0E+07
D-Glyceraldehdye 3-phosphate	1.3E+07	1.9E+07	9.6E+06	3.1E+06	5.2E+06	1.3E+07	2.1E+08	3.5E+08	2.8E+08
sn-Glycerol 3-phosphate	3.0E+07	1.0E+08	2.7E+07	1.2E+08	1.4E+08	5.2E+08	1.6E+09	1.9E+09	1.1E+09
Aconitate	1.1E+08	1.9E+08	1.2E+08	1.1E+08	1.5E+08	2.9E+08	1.4E+08	1.4E+08	1.2E+08
N-Acetylornithine	2.1E+08	2.5E+08	2.1E+08	4.0E+09	4.1E+09	4.0E+09	7.7E+07	8.4E+07	7.7E+07
Citrulline	1.5E+09	1.3E+09	1.4E+09	9.1E+06	6.3E+06	1.1E+07	2.2E+08	2.0E+08	2.3E+08
N-Carbamoyl-L-aspartate	9.2E+08	5.5E+08	7.3E+08	9.3E+06	1.5E+07	5.3E+07	1.9E+07	1.5E+07	1.8E+07
Gluconolactone	1.6E+07	4.7E+07	2.1E+07	1.6E+07	2.2E+07	6.5E+07	2.4E+07	2.1E+07	2.3E+07
Hydroxyphenylpyruvate	7.9E+06	5.4E+06	7.0E+06	3.8E+06	2.6E+06	2.3E+06	1.1E+06	1.1E+06	1.2E+06
Tyrosine	3.7E+08	8.2E+08	4.8E+08	3.5E+08	4.7E+08	1.1E+09	5.0E+08	4.9E+08	4.8E+08
Homovanillic acid (HVA)	1.1E+07	1.1E+07	1.0E+07	3.2E+07	3.8E+07	5.6E+07	2.4E+06	2.0E+06	3.6E+06
Homocysteic acid	3.1E+07	3.4E+07	3.9E+07	2.1E+07	2.1E+07	2.0E+07	3.8E+07	3.1E+07	3.5E+07

Table A-1. Continued.

COMPOUND	196_1	196_2	196_3	192_1	192_2	192_3	06-3228_1	06-3228_2	06-3228_3
3-Phosphoserine	4.9E+06	6.7E+06	6.2E+06	1.7E+07	1.1E+07	1.3E+07	2.6E+06	2.4E+06	1.9E+06
3-Phosphoglycerate	6.1E+08	6.7E+08	6.1E+08	6.6E+08	4.8E+08	4.5E+08	1.9E+09	2.2E+09	1.4E+09
Acetyllysine	9.4E+07	1.1E+08	1.2E+08	6.1E+08	5.6E+08	4.6E+08	1.1E+08	1.2E+08	1.0E+08
N-Acetylglutamate	3.2E+08	3.6E+08	3.7E+08	1.0E+08	1.4E+08	2.6E+08	1.7E+08	1.6E+08	2.0E+08
homocitrulline	3.5E+07	3.4E+07	3.7E+07	2.9E+07	2.9E+07	3.0E+07	1.5E+07	1.6E+07	1.5E+07
D-Gluconate	9.0E+07	1.7E+08	9.0E+07	4.6E+07	4.7E+07	1.1E+08	7.4E+09	8.1E+09	7.5E+09
D-Glucarate	1.0E+06	4.3E+06	1.4E+06	4.8E+05	1.1E+06	5.8E+06	1.5E+06	1.4E+06	1.3E+06
Jasmonate	4.5E+05	3.9E+05	5.9E+05	9.2E+05	1.0E+06	1.0E+06	3.7E+05	4.3E+05	3.4E+05
Deoxyribose phosphate	8.6E+07	8.1E+07	6.9E+07	2.9E+07	2.6E+07	2.7E+07	2.5E+08	3.5E+08	1.1E+08
Cystathionine	6.7E+07	5.4E+07	6.4E+07	1.8E+08	1.7E+08	1.3E+08	8.5E+06	8.6E+06	8.1E+06
deoxycytidine	6.9E+06	1.2E+07	9.6E+06	1.7E+06	2.6E+06	4.8E+06	9.0E+06	8.3E+06	9.3E+06
Ribose phosphate	5.6E+07	4.0E+07	2.8E+07	2.9E+07	3.7E+07	3.1E+07	6.6E+08	6.5E+08	5.8E+08
Uridine	1.5E+07	1.9E+07	1.6E+07	7.0E+06	1.1E+07	2.0E+07	2.7E+07	2.3E+07	2.6E+07
Shikimate-3-phosphate	2.5E+06	2.4E+06	2.8E+06	6.1E+04	1.2E+05	1.0E+05	2.0E+04	1.0E+04	0.0E+00
6-Phospho-D-gluconolactone	1.6E+06	9.8E+05	7.6E+05	5.0E+05	7.4E+05	2.8E+05	3.9E+07	4.0E+07	4.0E+07
Glucosamine phosphate	1.4E+07	1.0E+07	5.2E+06	2.0E+06	3.9E+06	1.7E+06	3.9E+07	4.4E+07	2.6E+07
S-Ribosyl-L-homocysteine	5.8E+06	6.4E+06	5.4E+06	6.4E+05	8.1E+05	1.5E+06	4.4E+06	4.3E+06	4.2E+06
6-Phospho-D-gluconate	8.9E+07	6.1E+07	3.9E+07	1.7E+07	2.4E+07	9.7E+06	1.6E+09	1.5E+09	1.3E+09
Xanthosine	2.8E+06	2.5E+06	3.2E+06	1.5E+07	1.5E+07	1.2E+07	9.3E+06	7.7E+06	1.1E+07
Ophthalmate	3.5E+05	1.2E+06	7.2E+05	6.3E+07	6.7E+07	6.9E+07	0.0E+00	0.0E+00	1.7E+04
Sedoheptulose 1/7-phosphate	9.4E+07	6.6E+07	4.7E+07	8.4E+07	6.0E+07	5.1E+07	3.1E+08	2.8E+08	1.4E+08
N-Acetylglucosamine 1/6-phosphate	1.2E+07	1.5E+07	1.1E+07	1.1E+07	1.1E+07	2.0E+07	3.7E+07	4.1E+07	3.0E+07
Glutathione	3.8E+06	9.8E+06	5.2E+06	1.1E+07	1.2E+07	2.3E+07	3.2E+06	3.0E+06	3.4E+06
IMP	3.8E+06	3.7E+06	3.1E+06	3.4E+06	6.1E+06	9.4E+06	3.9E+06	4.3E+06	3.6E+06
Trehalose 6-phosphate	6.5E+05	1.6E+06	8.9E+05	9.2E+06	1.5E+07	1.8E+07	1.3E+07	1.3E+07	1.1E+07
FMN	4.6E+06	4.8E+06	6.3E+06	3.7E+06	3.8E+06	3.8E+06	8.3E+06	8.6E+06	9.3E+06
UDP-glucose	1.8E+08	2.4E+08	2.5E+08	3.5E+07	3.9E+07	4.7E+07	1.3E+08	1.5E+08	1.3E+08
UDP-N-acetylglucosamine	5.5E+07	7.9E+07	8.4E+07	1.0E+08	4.8E+07	5.3E+07	5.9E+07	6.2E+07	5.7E+07
NAD+	1.8E+08	1.9E+08	1.9E+08	1.8E+08	1.8E+08	1.9E+08	1.4E+08	1.6E+08	1.4E+08
NADH	3.0E+06	3.9E+06	4.3E+06	1.7E+06	1.6E+06	1.3E+06	5.4E+06	7.5E+06	5.3E+06

Table A-1 Continued.

COMPOUND	196_1	196_2	196_3	192_1	192_2	192_3	06-3228_1	06-3228_2	06-3228_3
Mevalonate	1.7E+07	2.6E+07	2.0E+07	9.2E+06	1.5E+07	2.8E+07	1.2E+07	1.1E+07	1.4E+07
2-C-Methyl-D-erythritol 4-phosphate	1.8E+07	1.1E+07	2.0E+07	7.7E+06	7.5E+06	5.1E+06	1.6E+07	1.6E+07	1.4E+07
Isopentenyl diphosphate	5.3E+06	2.1E+06	4.4E+06	1.5E+06	2.7E+06	4.2E+06	1.8E+06	5.4E+06	1.4E+06
Geranyl diphosphate	4.2E+03	2.2E+03	4.5E+03	2.4E+03	0.0E+00	0.0E+00	1.9E+03	0.0E+00	0.0E+00
5-Phosphatomevalonate	3.6E+05	1.1E+05	1.0E+06	7.8E+05	6.2E+05	3.9E+05	9.0E+04	5.3E+05	2.0E+03

Table A-2. Staphylococcus Metabolites Normalized USA300, ST398, 132-96. This table lists the normalized intensities associated with metabolite annotations for strains USA300, ST398, and 132-96. USA300_1 had an optical density of 0.459, USA300_3 had an optical density of 0.489, USA300_3 had an optical density of 0.439, ST398_1 had an optical density of 0.427, ST398_2 had an optical density of 0.401, ST398_3 0.435, 132-96_1 had an optical density of 0.367, 132-96_2 had an optical density of 0.408, 132-96_3 had an optical density of 0.363.

COMPOUND	USA300_1	USA300_2	USA300_3	ST398_1	ST398_2	ST398_3	132-96_1	132-96_2	132-96_3
Dimethylglycine	1.4E+08	1.6E+08	1.6E+08	2.0E+07	3.0E+07	1.9E+07	4.8E+07	6.1E+07	6.5E+07
hydroxybutyrate	3.8E+08	3.6E+08	3.0E+08	9.4E+08	1.0E+09	9.8E+08	4.1E+08	6.1E+08	6.8E+08
Histamine	6.9E+05	6.9E+05	5.9E+05	4.7E+05	4.8E+05	5.4E+05	3.7E+04	3.5E+04	1.2E+04
Proline	2.9E+08	2.7E+08	2.5E+08	3.3E+08	3.3E+08	3.5E+08	6.1E+08	5.9E+08	4.4E+08
Fumarate	1.2E+07	1.3E+07	1.4E+07	1.6E+07	1.5E+07	1.9E+07	2.3E+07	2.4E+07	1.9E+07
2-Oxoisovalerate	9.9E+06	1.8E+07	1.7E+07	5.1E+06	5.1E+06	6.5E+06	2.9E+06	2.2E+06	1.7E+06
Valine/betaine	1.1E+09	7.6E+08	6.6E+08	5.1E+08	7.3E+08	5.4E+08	3.9E+08	6.5E+08	8.4E+08
Succinate/Methylmalonate	2.2E+09	1.5E+09	1.4E+09	2.0E+09	2.4E+09	1.9E+09	1.4E+09	2.2E+09	2.7E+09
3-Hydroxyisovalerate	1.5E+08	1.2E+08	1.0E+08	1.6E+08	1.8E+08	1.6E+08	2.2E+08	2.6E+08	2.9E+08
Homoserine/Threonine	4.9E+07	3.1E+07	3.1E+07	3.4E+08	3.9E+08	3.5E+08	2.2E+08	2.1E+08	1.5E+08
3-Methylthiopropionate	4.7E+06	3.9E+06	3.3E+06	5.2E+06	6.3E+06	5.0E+06	4.1E+05	1.4E+06	1.8E+06
Cysteine	5.8E+06	5.0E+06	4.0E+06	1.3E+06	1.6E+06	1.5E+06	6.8E+05	1.6E+06	2.6E+06
Citraconate	2.4E+07	2.1E+07	1.6E+07	1.6E+07	1.8E+07	1.9E+07	2.7E+07	3.5E+07	4.9E+07
N-Acetylputrescine	1.3E+06	6.6E+05	6.9E+05	1.4E+06	1.8E+06	1.5E+06	3.2E+06	3.7E+06	4.1E+06
Hydroxyproline	3.3E+07	2.2E+07	2.0E+07	1.7E+07	2.4E+07	1.7E+07	7.9E+07	8.1E+07	8.0E+07
Leucine/Isoleucine	6.3E+09	4.1E+09	3.7E+09	3.2E+09	4.5E+09	3.3E+09	2.9E+09	4.4E+09	6.1E+09
methyl succinic acid	1.4E+08	1.4E+08	1.4E+08	1.4E+08	1.7E+08	1.3E+08	6.6E+07	1.1E+08	1.0E+08
Asparagine	1.2E+09	1.1E+09	1.0E+09	1.3E+09	1.3E+09	1.3E+09	4.2E+08	5.9E+08	7.0E+08
Hydroxyisocaproic acid	7.8E+08	6.5E+08	6.2E+08	6.9E+08	7.9E+08	7.5E+08	9.8E+08	1.1E+09	1.1E+09
Ornithine	8.1E+07	9.2E+07	8.4E+07	1.7E+08	1.6E+08	1.8E+08	9.6E+07	9.0E+07	7.9E+07
Aspartate	4.8E+09	5.0E+09	4.6E+09	3.8E+09	3.9E+09	4.2E+09	2.4E+09	3.0E+09	3.4E+09
Homocysteine	8.4E+05	5.2E+05	4.2E+05	3.7E+05	4.9E+05	3.4E+05	9.6E+05	1.8E+06	2.2E+06
Anthranilate	8.0E+06	1.0E+07	1.0E+07	1.1E+07	1.1E+07	1.3E+07	1.1E+07	9.6E+06	8.4E+06
Hypoxanthine	2.0E+07	6.3E+06	5.7E+06	5.0E+06	1.2E+07	5.1E+06	1.1E+07	2.7E+07	4.1E+07
Salicylate	2.5E+07	3.8E+07	4.1E+07	3.4E+07	3.3E+07	3.9E+07	3.3E+07	2.9E+07	2.2E+07

Table A-2 Continued.

COMPOUND	USA300_1	USA300_2	USA300_3	ST398_1	ST398_2	ST398_3	132-96_1	132-96_2	132-96_3
Hydroxybenzoate	2.6E+07	2.8E+07	2.5E+07	3.1E+07	3.1E+07	3.0E+07	2.7E+07	2.5E+07	2.4E+07
Acetylphosphate	2.3E+08	2.6E+08	2.6E+08	1.7E+08	1.5E+08	2.2E+08	7.9E+07	9.1E+07	8.2E+07
alpha-Ketoglutarate	1.4E+08	1.9E+08	1.6E+08	8.3E+07	9.2E+07	1.0E+08	1.9E+08	3.0E+08	4.2E+08
Glutamine	2.0E+08	2.2E+08	2.1E+08	1.4E+08	1.5E+08	1.6E+08	3.9E+08	3.6E+08	3.5E+08
Lysine	6.5E+07	7.6E+07	7.0E+07	4.8E+07	4.7E+07	5.3E+07	3.7E+07	4.6E+07	5.2E+07
O-Acetyl-L-serine	3.8E+06	3.8E+06	2.0E+06	1.3E+06	6.4E+05	1.8E+06	1.5E+07	2.0E+07	1.4E+07
Glutamate	3.0E+09	3.1E+09	2.8E+09	2.8E+09	2.7E+09	3.0E+09	2.3E+09	2.8E+09	3.2E+09
Methionine	5.0E+08	3.5E+08	2.9E+08	3.2E+08	4.1E+08	3.6E+08	1.6E+08	3.1E+08	4.7E+08
Guanine	2.8E+07	9.1E+06	7.8E+06	7.6E+06	1.8E+07	8.3E+06	2.0E+07	4.6E+07	6.9E+07
Vanillin	4.0E+07	6.6E+07	6.1E+07	5.4E+07	5.0E+07	5.8E+07	7.5E+07	6.3E+07	6.4E+07
Xylitol	6.7E+06	3.2E+06	3.9E+06	4.0E+06	7.4E+06	3.8E+06	4.3E+06	9.4E+07	8.2E+06
Orotate	1.2E+08	5.4E+07	4.4E+07	6.1E+07	1.1E+08	5.5E+07	3.8E+07	8.3E+07	1.4E+08
Dihydroorotate	2.0E+07	8.9E+06	7.1E+06	1.9E+07	3.2E+07	1.3E+07	1.1E+07	2.5E+07	3.5E+07
pimelic acid	2.0E+07	2.4E+07	2.3E+07	2.5E+07	2.6E+07	2.5E+07	1.4E+07	1.9E+07	1.6E+07
Indole-3-carboxylate	1.4E+06	1.6E+06	1.5E+06	1.4E+06	1.5E+06	1.5E+06	1.2E+06	1.2E+06	1.1E+06
Phenylpyruvate	1.2E+08	1.9E+08	1.6E+08	1.3E+08	1.1E+08	1.4E+08	7.9E+07	7.4E+07	5.6E+07
Methionine sulfoxide	3.4E+07	1.5E+07	1.4E+07	1.1E+07	1.8E+07	9.9E+06	9.1E+06	1.9E+07	2.5E+07
Phenyllactic acid	1.5E+08	1.3E+08	1.2E+08	1.8E+08	2.2E+08	2.0E+08	1.7E+08	2.0E+08	2.0E+08
Cysteate	2.0E+08	2.6E+08	2.7E+08	1.6E+08	1.7E+08	1.6E+08	7.2E+07	7.7E+07	7.4E+07
Sulfolactate	1.7E+08	1.0E+08	7.4E+07	1.6E+08	2.1E+08	1.7E+08	5.2E+07	1.2E+08	1.8E+08
D-Glyceraldehdye 3-phosphate	2.1E+07	1.4E+07	1.8E+07	2.1E+07	3.9E+07	2.3E+07	5.2E+06	4.1E+07	5.7E+07
sn-Glycerol 3-phosphate	1.4E+08	5.2E+07	5.5E+07	1.6E+08	1.9E+08	1.3E+08	4.4E+07	1.4E+08	1.3E+08
Aconitate	1.5E+08	1.2E+08	1.0E+08	9.7E+07	1.2E+08	1.1E+08	1.6E+08	2.2E+08	2.8E+08
N-Acetylornithine	2.1E+08	1.8E+08	1.8E+08	3.8E+08	3.9E+08	3.9E+08	2.8E+09	2.7E+09	2.8E+09
Citrulline	3.8E+08	4.5E+08	4.1E+08	8.7E+08	8.1E+08	9.0E+08	2.9E+08	1.9E+08	8.9E+07
N-Carbamoyl-L-aspartate	5.0E+07	2.9E+07	3.2E+07	1.2E+08	1.2E+08	1.1E+08	5.1E+07	9.2E+07	1.1E+08
Gluconolactone	1.2E+08	8.1E+07	7.1E+07	3.3E+07	5.7E+07	3.3E+07	1.4E+07	3.8E+07	6.1E+07
Hydroxyphenylpyruvate	9.7E+05	1.6E+06	1.7E+06	1.5E+06	1.4E+06	1.8E+06	2.0E+06	1.9E+06	1.2E+06
Tyrosine	1.2E+09	8.6E+08	8.1E+08	6.3E+08	9.4E+08	6.7E+08	4.0E+08	7.6E+08	1.1E+09
Homovanillic acid (HVA)	8.3E+06	5.7E+06	4.3E+06	8.1E+06	1.0E+07	6.7E+06	1.1E+07	1.8E+07	2.4E+07

Table A-2 Continued.

COMPOUND	USA300_1	USA300_2	USA300_3	ST398_1	ST398_2	ST398_3	132-96_1	132-96_2	132-96_3
Homocysteic acid	2.3E+07	2.9E+07	3.3E+07	2.2E+07	2.1E+07	2.1E+07	1.0E+06	1.7E+06	2.4E+06
3-Phosphoserine	9.6E+06	8.9E+06	9.5E+06	5.0E+06	5.7E+06	5.5E+06	7.3E+06	8.3E+06	8.9E+06
3-Phosphoglycerate	1.9E+09	2.1E+09	1.7E+09	1.9E+09	1.6E+09	2.1E+09	4.6E+08	3.6E+08	3.3E+08
Acetyllysine	3.2E+08	2.8E+08	2.9E+08	1.7E+08	1.9E+08	1.8E+08	7.6E+07	7.9E+07	8.0E+07
N-Acetylglutamate	4.5E+08	4.4E+08	4.8E+08	4.6E+08	5.4E+08	5.3E+08	1.7E+08	2.6E+08	3.3E+08
homocitrulline	4.2E+07	4.0E+07	4.3E+07	4.2E+07	4.2E+07	4.2E+07	3.9E+07	3.8E+07	4.3E+07
D-Gluconate	1.3E+08	8.3E+07	8.6E+07	5.6E+07	8.4E+07	5.1E+07	6.3E+08	6.2E+08	5.4E+08
D-Glucarate	7.0E+06	2.3E+06	1.9E+06	1.6E+06	4.3E+06	1.7E+06	8.9E+05	3.9E+06	7.8E+06
Jasmonate	4.5E+06	8.8E+06	5.5E+06	4.1E+05	4.3E+05	5.5E+05	1.6E+06	1.7E+06	1.4E+06
Deoxyribose phosphate	1.5E+08	9.6E+07	1.2E+08	6.6E+07	6.9E+07	6.5E+07	1.9E+07	2.3E+07	2.3E+07
Cystathionine	1.4E+07	1.6E+07	1.7E+07	8.3E+06	8.1E+06	8.5E+06	1.0E+07	7.9E+06	7.3E+06
deoxycytidine	1.9E+07	1.5E+07	1.5E+07	2.1E+07	2.7E+07	2.4E+07	6.8E+06	1.1E+07	1.7E+07
Ribose phosphate	9.5E+06	7.9E+06	9.9E+06	6.0E+07	7.2E+07	6.5E+07	5.4E+07	8.6E+07	8.4E+07
Uridine	1.1E+07	6.6E+06	5.1E+06	5.3E+06	8.3E+06	5.4E+06	1.0E+07	1.7E+07	2.2E+07
Shikimate-3-phosphate	2.5E+04	3.4E+04	7.5E+03	2.8E+05	1.4E+05	4.1E+05	4.3E+05	3.2E+05	3.2E+05
6-Phospho-D-gluconolactone	1.9E+05	1.4E+05	1.6E+05	9.2E+05	8.1E+05	1.4E+06	6.3E+06	4.7E+06	3.8E+06
Glucosamine phosphate	1.2E+06	1.2E+06	1.6E+06	1.2E+07	1.1E+07	1.2E+07	1.2E+07	1.3E+07	1.2E+07
S-Ribosyl-L-homocysteine	3.8E+06	3.0E+06	2.8E+06	3.4E+06	4.3E+06	3.7E+06	2.6E+06	3.9E+06	4.6E+06
6-Phospho-D-gluconate	1.0E+07	1.0E+07	9.4E+06	3.1E+07	2.9E+07	4.6E+07	2.4E+08	2.6E+08	2.6E+08
Xanthosine	7.4E+06	5.7E+06	5.8E+06	9.3E+06	9.1E+06	9.0E+06	9.0E+06	9.3E+06	6.9E+06
Ophthalmate	1.7E+06	1.1E+05	5.2E+04	6.5E+04	9.9E+05	1.2E+05	1.3E+07	1.6E+07	2.1E+07
Sedoheptulose 1/7-phosphate	1.5E+07	1.5E+07	1.5E+07	5.9E+07	6.4E+07	6.3E+07	9.1E+07	1.1E+08	9.7E+07
N-Acetylglucosamine 1/6-phosphate	1.2E+07	6.6E+06	7.1E+06	1.2E+07	1.5E+07	1.3E+07	1.3E+07	1.3E+07	1.5E+07
Glutathione	5.9E+07	5.6E+07	5.4E+07	5.0E+07	6.2E+07	5.5E+07	1.2E+07	2.0E+07	2.6E+07
IMP	1.6E+05	1.3E+05	1.7E+05	4.2E+05	2.6E+05	4.2E+05	2.5E+06	1.8E+06	1.3E+06
Trehalose 6-phosphate	2.5E+07	2.4E+07	2.3E+07	4.2E+07	4.7E+07	4.4E+07	9.3E+06	1.3E+07	1.3E+07
FMN	2.9E+06	3.8E+06	3.7E+06	2.6E+06	3.1E+06	3.2E+06	2.5E+06	3.0E+06	3.1E+06
UDP-glucose	6.8E+07	6.3E+07	5.9E+07	7.6E+07	9.1E+07	8.5E+07	3.7E+07	4.1E+07	3.9E+07
UDP-N-acetylglucosamine	2.7E+07	2.1E+07	2.2E+07	7.5E+07	8.6E+07	8.9E+07	1.0E+08	4.9E+07	4.9E+07
NAD+	3.2E+08	3.6E+08	3.4E+08	2.8E+08	2.9E+08	3.3E+08	1.7E+08	1.8E+08	1.8E+08

Table A-2 Continued.

COMPOUND	USA300_1	USA300_2	USA300_3	ST398_1	ST398_2	ST398_3	132-96_1	132-96_2	132-96_3
NADH	3.1E+06	3.5E+06	3.8E+06	9.2E+06	1.1E+07	9.1E+06	1.6E+06	1.4E+06	1.1E+06
Mevalonate	1.8E+09	2.0E+09	1.5E+09	1.1E+09	1.3E+09	1.3E+09	1.4E+07	2.6E+07	4.5E+07
2-C-Methyl-D-erythritol 4-phosphate	9.1E+06	1.8E+07	1.9E+07	1.5E+07	9.1E+06	1.4E+07	8.2E+06	7.2E+06	5.7E+06
Isopentenyl diphosphate	2.5E+06	4.1E+06	4.1E+06	2.5E+06	7.8E+05	3.1E+06	2.1E+06	2.8E+06	1.9E+06
Geranyl diphosphate	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	4.7E+03	0.0E+00	2.4E+03
5-Phosphatomevalonate	8.6E+08	1.0E+09	9.7E+08	5.4E+08	5.7E+08	6.5E+08	7.5E+05	1.9E+05	4.0E+05

Table A-3. Staphylococcus Metabolites Normalized 182150, 182116, Blanks. This table lists the normalized intensities associated with metabolite annotations for strains 182150, 182116, and the blanks. 182150_1 had an optical density of 0.439, 182150_2 had an optical density of 0.409, 182150_3 had an optical density of 0.479, 182116_1 had an optical density of 0.331, 182116_2 had an optical density of 0.338, 182116_3 had an optical density of 0.350.

COMPOUND	182150_1	182150_2	182150_3	182116_1	182116_2	182116_3	MEDIA BLANK	EXTRACTION BLANK
Dimethylglycine	5.2E+07	6.8E+07	7.7E+07	7.6E+07	5.7E+07	8.0E+07	1.1E+06	6.5E+03
hydroxybutyrate	2.0E+08	2.1E+08	1.9E+08	2.2E+08	2.0E+08	2.2E+08	5.3E+07	4.6E+07
Histamine	3.1E+04	1.8E+04	8.9E+03	1.1E+04	1.7E+04	1.8E+04	0.0E+00	5.8E+02
Proline	1.4E+08	1.7E+08	1.4E+08	1.1E+08	9.1E+07	9.9E+07	4.1E+07	3.0E+03
Fumarate	6.8E+06	8.0E+06	9.9E+06	9.0E+06	9.1E+06	9.0E+06	3.1E+06	1.6E+06
2-Oxoisovalerate	1.9E+06	2.8E+06	2.1E+06	2.1E+06	1.7E+06	2.1E+06	5.3E+06	1.9E+05
Valine/betaine	7.1E+08	6.5E+08	6.9E+08	8.1E+08	5.8E+08	7.0E+08	3.6E+08	1.8E+04
Succinate/Methylmalonate	1.7E+09	1.8E+09	1.9E+09	2.2E+09	1.8E+09	1.9E+09	1.2E+09	2.6E+07
3-Hydroxyisovalerate	3.0E+08	3.5E+08	2.6E+08	3.2E+08	3.1E+08	2.8E+08	5.9E+07	8.4E+05
Homoserine/Threonine	4.6E+07	5.2E+07	4.6E+07	2.9E+07	2.9E+07	3.0E+07	9.5E+07	8.7E+04
3-Methylthiopropionate	1.8E+06	1.8E+06	1.7E+06	1.3E+06	1.4E+06	1.2E+06	4.0E+05	0.0E+00
Cysteine	1.6E+06	1.2E+06	1.0E+06	2.6E+06	7.9E+05	3.4E+05	1.3E+04	7.7E+02
Citraconate	1.7E+07	1.4E+07	1.7E+07	4.4E+07	2.6E+07	3.4E+07	6.4E+06	4.5E+05
N-Acetylputrescine	3.4E+05	3.2E+05	3.3E+05	7.7E+05	3.7E+05	5.2E+05	6.4E+05	0.0E+00
Hydroxyproline	3.5E+07	3.8E+07	3.1E+07	2.0E+07	1.6E+07	1.8E+07	8.1E+06	0.0E+00
Leucine/Isoleucine	3.9E+09	3.8E+09	4.0E+09	4.7E+09	3.1E+09	3.9E+09	2.0E+09	3.8E+04
methyl succinic acid	1.1E+08	1.3E+08	1.0E+08	1.3E+08	1.4E+08	1.4E+08	4.7E+07	2.0E+07
Asparagine	1.0E+09	1.1E+09	8.8E+08	1.1E+09	1.1E+09	1.1E+09	2.4E+08	1.3E+04
Hydroxyisocaproic acid	1.8E+09	2.1E+09	1.7E+09	2.1E+09	2.1E+09	2.0E+09	5.5E+07	7.5E+06
Ornithine	3.3E+08	3.6E+08	2.9E+08	1.5E+08	1.4E+08	1.6E+08	6.7E+06	1.7E+04
Aspartate	2.7E+09	2.5E+09	2.4E+09	2.1E+09	1.9E+09	2.0E+09	1.0E+09	1.4E+05
Homocysteine	4.4E+06	3.7E+06	3.1E+06	6.5E+06	6.2E+06	4.1E+06	7.1E+04	0.0E+00
Anthranilate	9.0E+06	1.0E+07	6.3E+06	5.1E+06	6.6E+06	6.0E+06	2.4E+06	1.7E+05
Hypoxanthine	1.3E+08	1.3E+08	1.8E+08	2.3E+08	1.1E+08	1.6E+08	1.5E+07	0.0E+00
Salicylate	3.3E+07	3.7E+07	2.6E+07	2.1E+07	3.2E+07	2.8E+07	7.2E+06	3.9E+06
Hydroxybenzoate	2.6E+07	2.8E+07	2.2E+07	2.0E+07	2.3E+07	2.2E+07	5.4E+06	3.5E+05

Table A-3 Continued.

COMPOUND	182150_1	182150_2	182150_3	182116_1	182116_2	182116_3	MEDIA BLANK	EXTRACTION BLANK
Acetylphosphate	1.8E+08	1.3E+08	1.5E+08	1.8E+08	1.6E+08	2.2E+08	7.0E+04	3.5E+04
alpha-Ketoglutarate	4.2E+07	5.2E+07	5.9E+07	1.0E+08	5.0E+07	6.4E+07	2.1E+07	1.0E+06
Glutamine	2.5E+08	2.7E+08	2.2E+08	2.6E+08	2.8E+08	2.7E+08	1.4E+07	7.1E+03
Lysine	2.8E+07	2.6E+07	2.6E+07	4.3E+07	3.4E+07	3.4E+07	1.5E+07	0.0E+00
O-Acetyl-L-serine	1.5E+07	1.8E+07	9.8E+06	2.4E+07	2.7E+07	2.7E+07	2.0E+03	2.4E+03
Glutamate	3.3E+09	2.9E+09	2.6E+09	2.3E+09	2.0E+09	2.1E+09	6.9E+08	5.5E+03
Methionine	1.9E+08	1.8E+08	2.4E+08	3.0E+08	1.5E+08	2.1E+08	1.5E+08	0.0E+00
Guanine	2.3E+07	2.2E+07	2.8E+07	3.7E+07	1.8E+07	2.5E+07	5.4E+06	0.0E+00
Vanillin	5.0E+08	2.5E+08	1.8E+08	1.5E+08	1.8E+08	1.4E+08	1.3E+08	4.1E+05
Xylitol	3.0E+06	3.6E+06	3.9E+06	4.3E+06	4.0E+06	3.9E+06	1.6E+07	2.1E+05
Orotate	1.2E+08	8.8E+07	5.7E+07	2.8E+07	1.9E+07	2.3E+07	1.9E+07	0.0E+00
Dihydroorotate	4.2E+08	4.4E+08	2.8E+08	9.6E+07	1.3E+08	1.3E+08	6.2E+06	0.0E+00
pimelic acid	1.9E+07	2.1E+07	1.5E+07	1.4E+07	2.3E+07	1.9E+07	7.2E+06	5.2E+06
Indole-3-carboxylate	1.2E+06	1.3E+06	1.1E+06	1.1E+06	1.3E+06	1.2E+06	3.8E+05	1.4E+05
Phenylpyruvate	8.1E+07	7.3E+07	5.4E+07	2.9E+07	5.5E+07	4.8E+07	1.1E+07	7.5E+05
Methionine sulfoxide	7.1E+06	8.7E+06	1.1E+07	1.6E+07	8.4E+06	1.3E+07	9.6E+06	1.9E+03
Phenyllactic acid	2.0E+08	2.4E+08	1.7E+08	2.0E+08	2.1E+08	2.0E+08	7.8E+06	9.5E+03
Cysteate	1.6E+08	1.6E+08	9.8E+07	5.5E+07	9.7E+07	8.7E+07	4.5E+06	1.6E+04
Sulfolactate	9.3E+07	9.4E+07	2.0E+08	3.4E+08	2.1E+08	2.4E+08	4.4E+07	3.6E+04
D-Glyceraldehdye 3-phosphate	1.7E+07	2.1E+07	6.5E+07	2.6E+08	2.4E+08	2.2E+08	1.1E+04	0.0E+00
sn-Glycerol 3-phosphate	3.8E+07	6.5E+07	6.2E+07	8.3E+07	7.6E+07	6.4E+07	4.5E+06	1.2E+04
Aconitate	9.9E+07	9.2E+07	1.1E+08	2.7E+08	1.6E+08	2.1E+08	3.3E+07	5.2E+06
N-Acetylornithine	1.8E+08	1.8E+08	1.1E+08	2.2E+08	2.0E+08	2.2E+08	4.6E+07	0.0E+00
Citrulline	1.8E+09	1.9E+09	1.5E+09	7.5E+08	7.8E+08	8.1E+08	1.0E+07	4.3E+04
N-Carbamoyl-L-aspartate	1.0E+09	1.0E+09	7.8E+08	2.1E+08	2.1E+08	2.3E+08	1.0E+07	0.0E+00
Gluconolactone	1.7E+07	1.9E+07	3.1E+07	4.3E+07	1.9E+07	2.8E+07	2.7E+07	5.2E+05
Hydroxyphenylpyruvate	2.1E+06	1.8E+06	1.6E+06	1.3E+06	1.5E+06	1.4E+06	1.2E+06	6.4E+05
Tyrosine	4.1E+08	4.3E+08	5.3E+08	6.7E+08	3.4E+08	4.8E+08	3.9E+08	7.5E+03
Homovanillic acid (HVA)	1.4E+07	1.5E+07	1.1E+07	1.1E+07	9.7E+06	9.9E+06	9.4E+05	3.8E+04
Homocysteic acid	3.6E+07	3.6E+07	2.1E+07	1.6E+07	2.0E+07	1.9E+07	1.2E+06	4.7E+03

Table A-3 Continued.

COMPOUND	182150_1	182150_2	182150_3	182116_1	182116_2	182116_3	MEDIA BLANK	EXTRACTION BLANK
3-Phosphoserine	4.4E+06	4.4E+06	4.8E+06	4.3E+06	3.2E+06	3.7E+06	2.0E+06	0.0E+00
3-Phosphoglycerate	8.1E+08	8.8E+08	8.2E+08	1.0E+09	9.5E+08	9.5E+08	5.8E+05	3.3E+03
Acetyllysine	9.9E+07	1.3E+08	8.1E+07	1.7E+08	1.6E+08	1.5E+08	6.0E+07	9.0E+02
N-Acetylglutamate	3.7E+08	4.8E+08	3.3E+08	3.4E+08	2.9E+08	3.2E+08	1.1E+08	3.6E+04
homocitrulline	3.7E+07	3.4E+07	2.8E+07	2.0E+07	2.2E+07	2.1E+07	2.6E+06	0.0E+00
D-Gluconate	2.9E+09	3.4E+09	7.1E+09	9.7E+09	9.0E+09	9.2E+09	9.4E+07	2.1E+05
D-Glucarate	1.4E+06	1.2E+06	3.5E+06	5.5E+06	1.7E+06	3.1E+06	1.1E+06	0.0E+00
Jasmonate	5.3E+05	5.0E+05	4.6E+05	4.5E+05	5.2E+05	5.1E+05	5.3E+05	3.7E+05
Deoxyribose phosphate	1.3E+08	1.5E+08	2.4E+08	2.1E+08	1.8E+08	1.7E+08	9.1E+05	0.0E+00
Cystathionine	5.7E+06	5.7E+06	5.1E+06	2.7E+06	2.9E+06	2.8E+06	3.0E+06	0.0E+00
deoxycytidine	1.3E+07	1.4E+07	9.7E+06	1.2E+07	9.8E+06	1.0E+07	1.7E+05	0.0E+00
Ribose phosphate	4.1E+07	5.8E+07	2.2E+08	5.8E+08	4.4E+08	4.5E+08	2.8E+06	0.0E+00
Uridine	2.5E+07	2.6E+07	2.2E+07	2.2E+07	2.1E+07	2.2E+07	2.8E+06	7.2E+02
Shikimate-3-phosphate	8.5E+05	5.7E+05	2.3E+06	1.3E+06	1.6E+06	1.5E+06	0.0E+00	0.0E+00
6-Phospho-D-gluconolactone	2.9E+06	3.0E+06	3.4E+07	6.6E+07	5.6E+07	5.6E+07	0.0E+00	0.0E+00
Glucosamine phosphate	5.4E+06	8.6E+06	2.5E+07	3.7E+07	3.3E+07	3.1E+07	9.8E+02	0.0E+00
S-Ribosyl-L-homocysteine	1.5E+07	1.6E+07	1.1E+07	2.2E+07	2.6E+07	2.6E+07	2.9E+05	0.0E+00
6-Phospho-D-gluconate	1.4E+08	1.3E+08	1.7E+09	2.9E+09	2.5E+09	2.5E+09	9.4E+03	2.3E+03
Xanthosine	8.3E+06	8.4E+06	8.5E+06	2.9E+06	3.2E+06	3.1E+06	4.7E+06	0.0E+00
Ophthalmate	1.6E+05	1.6E+05	8.9E+04	9.1E+05	1.2E+05	1.7E+05	1.3E+04	0.0E+00
Sedoheptulose 1/7-phosphate	2.5E+07	4.3E+07	1.6E+08	2.0E+08	1.5E+08	1.5E+08	4.4E+05	2.7E+03
N-Acetylglucosamine 1/6-phosphate	1.0E+07	1.4E+07	1.5E+07	2.0E+07	1.8E+07	1.9E+07	3.3E+06	0.0E+00
Glutathione	1.3E+07	1.1E+07	1.1E+07	1.1E+07	6.4E+06	5.2E+06	3.1E+06	0.0E+00
IMP	1.6E+06	7.9E+05	2.5E+06	1.4E+06	1.4E+06	5.4E+06	7.7E+04	0.0E+00
Trehalose 6-phosphate	6.6E+05	6.5E+05	1.1E+06	1.4E+06	4.9E+05	8.8E+05	1.9E+06	0.0E+00
FMN	3.3E+06	3.9E+06	2.4E+06	4.6E+06	5.5E+06	5.0E+06	6.5E+03	0.0E+00
UDP-glucose	2.5E+08	3.3E+08	3.5E+08	6.2E+08	4.6E+08	4.4E+08	3.1E+04	5.3E+04
UDP-N-acetylglucosamine	5.1E+07	7.6E+07	4.4E+07	5.0E+07	5.6E+07	4.9E+07	1.8E+06	2.6E+03

Table A-3 Continued.

COMPOUND	182150_1	182150_2	182150_3	182116_1	182116_2	182116_3	MEDIA BLANK	EXTRACTION BLANK
NAD+	2.3E+08	2.4E+08	1.7E+08	1.9E+08	2.0E+08	2.0E+08	3.7E+03	0.0E+00
NADH	5.8E+06	7.2E+06	4.6E+06	5.4E+06	4.9E+06	4.2E+06	0.0E+00	0.0E+00
Mevalonate	2.4E+07	2.5E+07	2.5E+07	3.0E+07	2.2E+07	2.5E+07	1.0E+07	6.7E+04
2-C-Methyl-D-erythritol 4-phosphate	1.7E+07	1.9E+07	1.3E+07	1.0E+07	1.7E+07	1.1E+07	9.6E+04	2.9E+04
Isopentenyl diphosphate	1.4E+06	3.5E+06	3.9E+06	1.8E+06	3.5E+06	2.2E+06	9.1E+05	1.7E+05
Geranyl diphosphate	8.8E+03	2.3E+03	0.0E+00	2.3E+03	0.0E+00	0.0E+00	0.0E+00	0.0E+00
5-Phosphatomevalonate	2.1E+05	1.1E+06	3.1E+05	1.4E+05	2.3E+05	9.3E+05	4.1E+03	2.7E+03

Vita

Matthew Keller was born in California and raised in Connecticut and Tennessee. He graduated Summa Cum Laude from the University of Tennessee, Chattanooga with a B.S. in Chemistry. Academic interests include mass spectrometry, analytical separations, and molecular biology. Matthew got married in 2019 to his best friend, Rose. They currently live in Knoxville.