ANALYSIS OF COMPLEX SAMPLES BY MASS SPECTROMETRY LEADS TO INSIGHTS INTO SYSTEM DYNAMICS

by

Jesse Thomas Peach

A dissertation submitted in partial fulfillment

of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY

Bozeman, Montana

April 2021

DEDICATION

To my lovely children Aysa and Torin, for being constant sources of strength and inspiration.

ACKNOWLEDGEMENTS

Thank you first and foremost to my advisor Dr. Brian Bothner for his patience, guidance and tireless work on my behalf during my graduate tenure at Montana State. Simply put, I would not be at this point without his mentorship, for which I am eternally grateful. I would also like to thank my graduate committee of Dr. Brent Peyton, Dr. Mary Miles and Dr. Mark Young for being exemplary in every way. I was also blessed with amazing collaborators including Dr. Rebecca Mueller, Dr. Dana Skorupa, Dr. Margaux Mesle and Stephanie Wilson. My journey would not have been possible without the help of the members of the Bothner Lab Group. Thank you, Bothner Lab members, past and present, for your collective knowledge and enthusiasm.

Although I had wonderful academic mentors and peers, my family makes up the base of the pillars that supported me during the last 5 years. Megan, thank you for letting me chase my dreams even though it made your life more challenging. Your unyielding trust in me over the course of this graduate experience has truly meant the world to me. Aysa you are my muse and make each and every day brighter and more vibrant. Thank you for being who you are, because who you are will always be more than enough. Torin, you were born in the middle of this PhD process and although I am very proud to obtain my doctorate, you are the true treasure that I gained over the last 5 years. Your energy and big heart always make me smile and feel thankful to have you in my life. I love you both to the end of all things. Thank you to my mom and dad for their continued support and for always valuing education. Thank you to my sisters KC and Rikki as well for being able to make me laugh, especially when I need it most. Lastly, thank you to my mother-in-law Phyllis for her help and kindness. I love you all and would not be here without each and every one of you.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

ABSTRACT

Systems biology offers a holistic approach to biological science. In its most complete form, systems biology requires comprehensive data encompassing all of the parts or molecules across a set of hierarchical networks. To obtain and analyze the comprehensive and large datasets required for systems biology analysis, biologists have taken advantage of new technology and computational tools. Over the last few decades, advances in computational modeling and analysis technology has dramatically increased the efficacy of systems biology and the understanding of the natural world. However, systems biology is still an emerging discipline. The overwhelming scale of potential biological data that has yet to be described, coupled with interpretation and application obstacles, leaves much work to be accomplished. One aspect of systems biology that needs development is the interpretation and analysis of temporal biological data. Temporal data reveals more about biological phenomena than static data as biology is inherently dynamic. This dissertation explores the benefits of temporal profiling of complex samples to make time-resolved conclusions about complicated biological questions. Three research projects are the backbone of this document, with a chapter being devoted to each. Chapter 2 describes the development of a comprehensive method for extraction and mass spectrometry analysis of several different fractions from hot spring sediment. Chapter 3 delves into a multi-omics analysis tracking changes over the course of three years in a thermoalkaline spring system in Yellowstone National Park. It defines how specific extracellular small molecules correlate with microbial fitness. Specifically, how unique nitrogen and sulfur containing molecules in the sediment drive archaeal abundance and diversity. The final chapter introduces the concept of a ‘dynamic biosignature’, a set of metabolites that have similar responses to known biomarkers, in this case pro-inflammatory cytokines. A cohort of metabolites was identified that provided mechanistic insight into the inflammatory response. Overall, this dissertation provides examples of systems biology analysis and provides evidence that static, single time-point datasets fail to capture that which is the essence of biology - change.

CHAPTER ONE

INTRODUCTION TO DISSERTATION

At the dawn of the 21st century, the Human Genome Project inspired a new generation of scientists and provided the platform to create a novel approach for the study of life that became known as modern systems biology. This new methodology of exploring biology arose due to the difficulty in obtaining genomic data and the challenges presented with whole genome analysis (Ideker, Chuang). Essentially, new technology and computational modeling methods were needed that could collect and disseminate large, comprehensive datasets that contained all the parts in a biological system (Hood). The integration of biology with emerging technologies and advanced computer modeling revealed a new biological world in not just genomics, but in other biological systems as well (Villa).

Three other biological research areas that have greatly benefitted from the development of systems biology are transcriptomics, proteomics and metabolomics. Transcriptomics involves RNA, proteomics focuses on protein while metabolomics concentrates on small molecules found in organisms. Together these three levels of biological analysis, along with genomics, make up the omics cascade (Dettmer). While interconnected, each level shows unique information about the biology of an organism. Genomics indicates what is biologically possible, transcriptomics signifies what seems to be occurring, proteomics shows the machinery behind the action and metabolomics is the closest to a true phenotypic description (Baumann). While data from each portion of the omics cascade can be illuminating, combining these datasets shows the true power of systems biology (Buescher, Cao, Rawle).

Although there have been several technological and computational advances in the last several decades that have led to the evolution of systems biology, this dissertation will focus on the use of mass spectrometry (MS) technology and temporal computational modeling. MS can be used to measure the mass and quantify nearly all classes of biomolecules (Dettmer, Domon, Patti). It has become an invaluable tool for systems biology due to its ability to generate robust, quantifiable datasets of entire biological systems in a rapid and efficient manner (Feng). A challenge with biological samples is that they are usually highly complex and isolating a specific set of biological data can be difficult. MS front end advances have greatly helped this issue with the development of superior chromatographic separation techniques including column technology advancements and chromatography system improvements (Lei, Ren). Despite the significant improvement in sensitivity, speed, and accuracy of MS instrumentation in the last 10 years, sample preparation and extraction methods remain a critical part of obtaining high quality omics data (Raftery). Chapter two showcases an innovative pipeline for the obtaining multiple classes of biomolecules from a complex sample. This method produces three different classes of molecules: purified and digested protein, extracellular small molecules, and intracellular small molecules or metabolites from hot springs sediment. The multi-step process is involved and requires individual extraction techniques for each fraction along with specialized liquid chromatography mass spectrometry (LCMS) conditions yet provides a comprehensive dataset for thermal sediment.

Temporal, or time-dependent, computational techniques are explored in Chapters Three and Four. Chapter Three includes the use of the methodology from Chapter Two to analyze longitudinal MS data over three years in a meta-analysis, coupled with geochemical and

metagenomic data, from thermoalkaline hot springs in Yellowstone National Park (YNP). An untargeted approach is taken wherein differences between the entire datasets are investigated over the course of the study. From this examination, global changes were seen in some datasets over the course of the study, like the small molecule data. Untargeted analysis can also isolate specific differences between datasets, as seen in a cohort of nitrogen and sulfur containing compounds that are modulated over the course of the study. This long-term, longitudinal study was able to resolve the loss of archaeal fitness to a decline in unique nitrogen and sulfur containing compounds, most likely due to snow-pack variation and groundwater mixing.

Annual studies like the multi-year study of Yellowstone thermal features in Chapter Three are rare due to the logistical challenges they present. However, not all temporal studies need to span a long time frame to reveal important biological characteristics. This is seen in Chapter Four, where serum samples are collected over a 4-hour period and still provide temporal data in a short amount of time. Serum metabolites were explored using LCMS and initial results did not provide evidence of metabolic differences between groups. However, using the pattern of response and response values, subtle differences were isolated between the cohort of participants that provided mechanistic insight into the inflammation response.

Biology is inherently dynamic; therefore, the analysis of biological data should be dynamic as well. Static modeling of complex, changing systems can still reveal answers to biological questions, but may miss subtle variations that can would otherwise not be elucidated (Richelle, Zupanic). The true power of systems biology is not just to differentiate groups, but to differentiate time-resolved modalities. This dissertation explores this concept with several

different examples showing the capacity of the union of systems biology data and temporal computation.

Literature Cited

BAUMANN, S., KALKHOF, S., HACKERMULLER, J., OTTO, W., TOMM, J. M., WISSENBACH, D. K., ROLLE-KAMPCZYK, U. & VON BERGEN, M. 2013. Requirements and Perspectives for Integrating Metabolomics with other Omics Data. *Current Metabolomics,* 1**,** 15-27.

BUESCHER, J. M. & DRIGGERS, E. M. 2016. Integration of omics: more than the sum of its parts. *Cancer & metabolism,* 4**,** 4-4.

CAO, H. X., ZHANG, A. H., SUN, H., ZHOU, X. H., GUAN, Y., LIU, Q., KONG, L. & WANG, X. J. 2015. Metabolomics-proteomics profiles delineate metabolic changes in kidney fibrosis disease. *Proteomics,* 15**,** 3699-3710.

CHUANG, H.-Y., HOFREE, M. & IDEKER, T. 2010. A decade of systems biology. *Annual review of cell and developmental biology,* 26**,** 721-744.

DETTMER, K., ARONOV, P. A. & HAMMOCK, B. D. 2007a. Mass spectrometry-based metabolomics. *Mass Spectrom Rev,* 26**,** 51-78.

DETTMER, K., ARONOV, P. A. & HAMMOCK, B. D. 2007b. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews,* 26**,** 51-78.

DOMON, B. & AEBERSOLD, R. 2006. Review - Mass spectrometry and protein analysis. *Science,* 312**,** 212-217.

FENG, X. J., LIU, X., LUO, Q. M. & LIU, B. F. 2008. MASS SPECTROMETRY IN SYSTEMS BIOLOGY: AN OVERVIEW. *Mass Spectrometry Reviews,* 27**,** 635-660.

HOOD, L. 2003. Systems biology: integrating technology, biology, and computation. *Mechanisms of Ageing and Development,* 124**,** 9-16.

IDEKER, T., GALITSKI, T. & HOOD, L. 2001. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet,* 2**,** 343-72.

LEI, Z. T., HUHMAN, D. V. & SUMNER, L. W. 2011. Mass Spectrometry Strategies in Metabolomics. *Journal of Biological Chemistry,* 286**,** 25435-25442.

PATTI, G. J., YANES, O. & SIUZDAK, G. 2012. Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology,* 13**,** 263-269.

RAFTERY, D. 2016. *Mass Spectrometry in Metabolomics*, Springer.

RAWLE, R. A., HAMERLY, T., TRIPET, B. P., GIANNONE, R. J., WURCH, L., HETTICH, R. L., PODAR, M., COPIÉ, V. & BOTHNER, B. 2017. Multi-omics analysis provides insight to the Ignicoccus hospitalis-Nanoarchaeum equitans association. *Biochimica et biophysica acta. General subjects,* 1861**,** 2218-2227.

REN, J. L., ZHANG, A. H., KONG, L. & WANG, X. J. 2018. Advances in mass spectrometry-based metabolomics for investigation of metabolites. *Rsc Advances,* 8**,** 22335-22350.

RICHELLE, A., DAVID, B., DEMAEGD, D., DEWERCHIN, M., KINET, R., MORREALE, A., PORTELA, R., ZUNE, Q. & VON STOSCH, M. 2020. Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. *npj Systems Biology and Applications,* 6**,** 6.

VILLA, A. & SONIS, S. T. 2020. 2 - System biology. *In:* SONIS, S. T. & VILLA, A. (eds.) *Translational Systems Medicine and Oral Disease.* Academic Press.

ZUPANIC, A., BERNSTEIN, H. C. & HEILAND, I. 2020. Systems biology: current status and challenges. *Cellular and Molecular Life Sciences,* Feb**,** 379-380.

CHAPTER TWO

OPTIMIZATION OF THERMAL SMALL MOLECULE AND PROTEIN MASS SPECTROMETRY ANALYSIS

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Jesse T. Peach

Contributions: Conceived the study, performed the extractions, MS and SDS-PAGE analysis, interpreted the results and wrote the manuscript

Co-Author: Sutton Kanta

Contributions: Performed the extractions and SDS-PAGE analysis and interpreted the results

Co-Author: James Larson

Contributions: Performed the SDS-PAGE analysis

Author: Eric Boltinghouse

Contributions: Performed the extractions

Co-Author: Rebecca Mueller

Contributions: Conceived the study

Co-Author: Ganesh Balasubramanian

Contributions: Conceived the study and performed the MS analysis

Author: Mohammed Refai

Contributions: Conceived the study and performed the extractions

Co-Author: Brent M. Peyton

Contributions: Conceived the study, and assisted in the manuscript

Co-Author: Brian Bothner

Contributions: Conceived the study, interpreted the results and assisted in the manuscript

Manuscript Information

Jesse T Peach, James Larson, Sutton Kanta, Eric Boltinghouse, Rebecca Mueller, Ganesh Balasubramanian, Mohammed Refai, Brent Peyton and Brian Bothner

Analytical Biochemistry

Status of Manuscript:
_X_ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Elsevier

Abstract

ABSTRACT- Thermal springs are understudied extreme environments that contain microorganisms with novel metabolic pathways and enzymes. These pathways and enzymes have the potential to impact science and industry, largely due to the promise of thermostable proteins. A major challenge associated with thermal spring analysis has been generating high quality mass spectrometry data. To the best of our knowledge, optimization of thermal specific mass spectrometry data has not been described. Therefore, an extensive optimization of methods was undertaken to analyze extracellular small molecule, intracellular small molecule and protein fractions from sediment slurry samples gathered at several thermal sites in Yellowstone National Park (YNP). A comprehensive pipeline was constructed including extraction methods and liquid chromatography mass spectrometry (LCMS) conditions for each fraction.

Introduction

The use of mass spectrometry (MS) in the analysis of microorganisms has greatly increased our understanding of microbial life, metabolism and the world in which they live (Sanguinetti). However, current techniques focus on mesophilic organisms and may not translate effectively to groups of other microbes, such as thermophiles. This is particularly concerning when analyzing protein, as thermophilic proteins are enhanced through the substitution of amino acids to promote thermal stability (Gromiha). As such, MS methods for thermophilic microbe analysis need to be developed to generate more accurate and robust MS datasets.

MS analysis of microbes and their habitat can be used to uncover a wide range of microbial data including protein expression and modification, upregulated metabolites and

environmental small molecules with specific metabolic functions (van der Meer). Comprehensive datasets can then reflect a global view of microbial communities and their interactions with their environment. Investigating thermophiles could have important industrial and scientific impacts (Wingen, DeCastro). This is due to the presence of novel thermostable protein or unique metabolism that has yet to be described (Barns, Hedulund, Mueller). In this paper, we seek to construct an extensive methodology for MS specific analysis of thermophilic microorganisms and their environment.

Materials and Methods

Sample collection

Samples were collected at the Five Sisters springs in White Creek Drainage in Yellowstone National Park (**Figure 2.1**). 15mL samples of sediment slurry, composed of ~7mL sediment and ~8mL water, were collected sterile Falcon tubes in triplicate from each thermal site. Upon collection, each sample was immediately frozen in a dry ice, ethanol bath and placed in a cooler containing dry ice. After returning to the lab, samples were placed into long-term storage in a -80°C freezer. When ready to analyze, the samples were examined via an extensive pipeline, shown in **Figure 2.2**. Completion of the pipeline resulted in the generation of three different types of MS data: extracellular and intracellular small molecules and proteomics data.

Cell extraction

Once ready to extract, samples were removed from the -80°C freezer and thawed in a 40°C water bath. When thawed, samples were immediately placed on ice. For the remainder of the cell extraction procedure, samples were kept on ice whenever possible. Samples were then

Figure 2.1. The Five Sisters hot springs in Yellowstone National Park. This site consists of five thermoalkaline springs with variable connectivity between the springs.

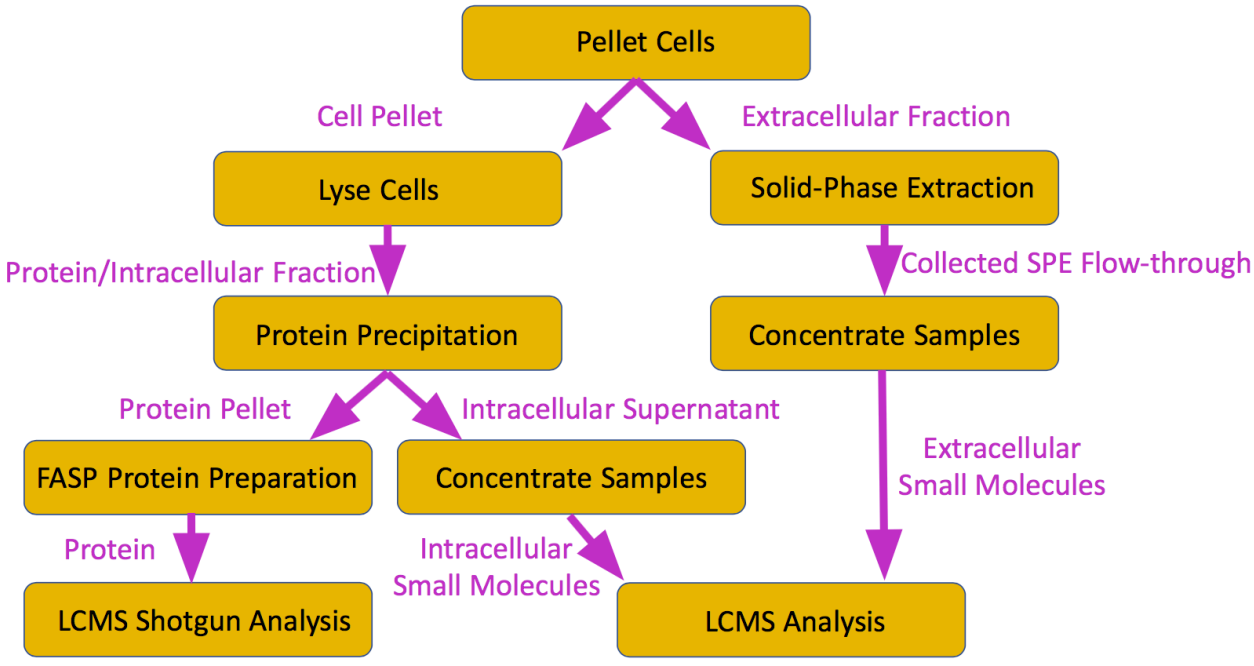


Figure 2.2. Workflow from pelleted cells to generate extracellular small molecules, intracellular small molecules and proteomics data.

agitated on a vortex machine vigorously at the maximum setting for one minute. After a minute wait period, the process was completed a second time to loosen microbes embedded in the sediment. Thawed and agitated sediment slurry samples were then gently spun at 400 RPM for four minutes at -4°C to pellet the sediment. Some samples with finer sediment required centrifuging a second to create a clear supernatant. After centrifugation, the supernatant was removed and placed in clean, pre-labeled tubes. The sediment was then washed with 5mL of Milli-Q water and the vortex and centrifugation process was repeated. The wash supernatant was collected and added to the previously collected supernatant. Samples were then centrifuged at 10,000 RPM for 15 minutes at -4°C to create a cell pellet. At this point, the supernatant was collected for environmental solid phase extraction (SPE) analysis and enough 0.01M phosphate buffer solution (1X PBS) was added to the original tube to cover the cell pellet.

Environmental extracellular small molecule SPE extraction

SPE extraction and concentration has been shown to generate robust and sensitive data from aqueous samples via several environmental systems (Dittmar, Faraji). A modified SPE method from Spencer, et al., 2014 was used to analyze the extracellular small molecule fraction. Briefly, the environmental small molecule layer was first acidified to pH 2 using formic acid (Fisher Chemical, Hampton, NH). 3mL Agilent Bond Elut PPL SPE cartridges (Agilent Technologies, Santa Clara, CA) connected to a Biotage VacMaster SPE Manifold (Biotage, Uppsala, Sweden) under a vacuum were used for the analysis. Cartridges were washed with two cartridge volume additions of MeOH (Fisher Chemical, Hampton, NH) followed by rinses with two cartridges of Milli-Q water and an additional cartridge of MeOH. A final rinse of two volumes of pH2 acidified Milli-Q water prepared the cartridges for analysis. The acidified

samples were passed through the cartridges and rinsed with two cartridge volumes of acidified Milli-Q. The cartridges were held under a nitrogen stream until the packing material was completely dry. Environmental extracellular small molecules were then eluted with the addition of 1mL of MeOH into a clean vial. Following elution, samples were further concentrated under negative pressure using a Concentrator Plus (Eppendorf, Hamburg, Germany) until completely dry and then stored in a -80°C freezer until LCMS analysis.

Cell lysis

Cell pellets in PBS were gently pipetted up and down several times before centrifuging for four minutes at 400 RPM at -4°C. PBS supernatant was removed and two volumes of extraction buffer was added to the cell pellet consisting of 8M urea (Fisher Chemical, Hampton, NH), 0.1M Tris-HCL (MilliporeSigma, Munich, Germany), 50mM ethylenediaminetetraacetic acid (EDTA) (MilliporeSigma, Munich, Germany) and 1X protease inhibitor mix (MilliporeSigma, Munich, Germany). Three rounds of a freeze/thaw procedure were then completed using liquid nitrogen. This procedure consisted of placing samples in liquid nitrogen for 10 seconds, or until frozen, and then allowing them to thaw at room temperature. Following the freeze/thaw, cell pellets in the extraction buffer were placed in a beaker with -80°C zinc beads. Samples were then sonicated using a Biologics Inc. Ultrasonic Homogenizer 3000 (Biologics, Manassas, VA) using a 40% duty cycle for three minutes. This process was repeated a second time, which increased the cell lysis and intracellular small molecule recovery.

Intracellular small molecule and protein recovery

After cell lysis, samples were centrifuged at 15,000 RPM for 30 minutes at -4°C to pellet cell debris. The intracellular small molecule and protein rich supernatant was removed and

placed in a clean vial while the cell debris pellet was washed with an additional 50µL of extraction buffer. The cellular debris in extraction buffer was vortexed for an additional minute followed by an additional round of centrifugation at 15,000 RPM for 30 minutes at -4°C. The supernatant wash was collected and added to the protein and intracellular small molecule rich vial. A protein precipitation was then performed with the addition of four volumes of ice-cold acetone (Fisher Chemical, Hampton, NH) at -20°C after which samples were placed in a -80°C freezer for two hours. Samples were spun for five minutes at 5,000 RPM at -4°C and the resulting metabolite supernatant was removed and placed in a clean, pre-labeled vial. Buffer was placed in the original vial to cover the protein pellet as described in the next section. Metabolite samples were then concentrated in a Speed Vac to dryness and stored in a -80°C to await LCMS analysis.

Filter-aided sample preparation of protein

The protein pellet was immediately resuspended in buffer containing 4M urea, 0.1M ammonium bicarbonate (ABC) (MilliporeSigma, Munich, Germany), 0.1M dithiothreitol (DTT) (Fisher Chemical, Hampton, NH) and Milli-Q water to two times the volume of the pellet. The sample was incubated in the buffer at room temperature for one hour. For this procedure, Agilent Bond Elut 3K Centrifuge Device Columns (Agilent Technologies, Santa Clara, CA) were used to extract and purify protein. Columns were prepared with the addition of 200µL Milli-Q water followed by ten minutes of centrifugation at 5,000 RPM. Continual checks of the column were performed throughout the procedure to verify that the column did not completely dry out. After centrifugation, 100µg of soluble protein was added, generally around 50µL, and the column was centrifuged at 8,000 RPM for ten minutes This process was repeated, and the flow through was

discarded both times as needed so the vial did not overflow. Centrifugation was followed by the addition of 100µL of iodoacetamide (Fisher Chemical, Hampton, NH) and a 25-minute incubation in the dark at room temperature. After 25 minutes, 100µL of 10mM DTT was added to quench the remaining iodoacetamide and the samples were allowed to sit at room for an additional ten minutes. Columns were then centrifuged for ten minutes at 8,000 RPM and the flow through was again discarded as needed. Four rounds of ABC washes were next completed. Each wash consisted of the addition of 200µL of ABC followed by centrifugation at 8,000 RPM for ten minutes. After the washing procedure was completed the flow through was discarded.

Protein digestion

Thermolysin (Promega, Madison, WI) was obtained and placed in a thermolysin digestion buffer containing 50mM Tris-HCL, and 0.5mM $CaCl_2$ (Fisher Chemical, Hampton, NH). Thermolysin was added to the buffer at a concentration of 1mg/mL. 50µL of thermolysin buffer at pH 6 was then added to the spin column along with 5uL of suspended thermolysin. Protein solution pH, incubation temperature and incubation time were all tested to determine the optimal digestion conditions. pH conditions varied from pH 6-8, while temperatures tested ranged from 65°C-85°C and incubation times ranged from 15 minutes to four hours. Digestion effectiveness was analyzed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) to determine the effectiveness of the digestion. After digestion, the columns were centrifuged for 30 minutes at 8,000 RPM and the flow through was collected in a clean vial. The collected peptide solution was then concentrated in a Speed Vac to dryness. Samples were resuspended in 40µL of a solution composed of 0.1% formic acid and 0.05% HPLC grade

acetonitrile (ACN) (Fisher Chemical, Hampton, NH) in HPLC grade water (Fisher Chemical, Hampton, NH) and transferred to a MS vial to be stored at -4°C until ready for LCMS analysis.

LCMS analysis

All LCMS analysis was completed at the Proteomics, Metabolomics and Mass Spectrometry facility at Montana State University. The small molecule fractions were analyzed on an Agilent 6538 Q-TOF MS (Agilent Technologies, Santa Clara, CA) coupled to an Agilent 1290 UHPLC (Agilent Technologies, Santa Clara, CA) using hydrophilic interaction chromatography (HILIC). For HILIC separation, a Cogent Diamond Hydride HPLC column (Microsolv, Greater Wilmington, NC) was utilized along with mobile phases consisting of water with 0.1% formic acid (A) and ACN with 0.1% formic acid (B). A 15-minute run time was used, starting with 100% of mobile phase B and moving to 30% B in a linear gradient over 14 minutes. At 14 minutes, mobile phase B was increased back to 100% for the final minute of the LC run. Flow was kept at 600µL/minute and the column compartment temperature was set at 30ºC. For the MS component, electrospray ionization was used in positive mode. After LCMS analysis, data was converted to mzML format with MSConvert (Chambers) and then coordinated using mzMine (Pluskal) with an intensity cutoff of 1,000 and a duplicate ion retention time window of 0.1 seconds. Blank samples were also analyzed and a cutoff of 5:1, sample to blank, was used to determine sample background and machine noise. Data analysis was completed using MetaboAnalyst and R code (Kew, Rohart).

Proteomic samples were separated via a Dionex nano-ultra-high performance liquid chromatography MS (Thermo Fisher Scientific, Waltham, MA) using reverse-phase chromatography. A Bischoff ProtoSil Eurobond C-18 column (Bischoff Chromatography,

Munich, Germany) was used in a 130-minute method. Mobile solvent (A) was water with 0.1% formic acid and (B) was acetonitrile with 0.1% formic acid. Mobile gradient started at 95% A from 0 to 2-minutes. From 2 to 122-minutes, B was ramped up to 30%, after which it was again increased to 80% and held until 125 minutes. Solvent concentrations then returned to the starting concentrations for the remainder of the analysis. Flow rate was kept constant at 0.4µL/minute. Samples were then analyzed with a Bruker MaXis Impact MS (Bruker, Bilerica, MA) using electrospray ionization with runs being completed in positive mode. Raw data was converted with MSConvert and SearchGUI (Barsnes) was used to search against proteomes from a compiled list of FASTA files from UniProt (The Uniprot Consortium) including a cohort of bacteria and archaea. Data analysis was completed using PeptideShaker (Vaudel). For both SearchGui and PeptideShaker, a precursor and product error of 35ppm was used along with a maximum of three missed cleavages.

Results and Discussion

The goal of this analysis pipeline was to increase the number of detected small molecules and peptides with sediment extraction and LCMS optimization. For the extracellular small molecules, the use of the SPE manifold was instrumental in improving coverage. Although the small molecules in the thermal environment were very diverse, they appeared to be in low concentration. However, the use of the PALL SPE cartridges and a SPE manifold concentrated the environmental samples and worked excellently with LCMS analysis. Initial unique small molecule counts without SPE averaged slightly less than 400 features per sample with an intensity cutoff of 1,000. With SPE cartridges, the number of features increased to over 1,700 on average from the same hot spring samples.

After data processing, extracellular data examination is possible in several different ways. Global profiling can be completed to determine the relationships between springs in the form of 2-D principal component analysis (PCA) score plots using MetaboAnalyst (**Figure 2.3A**) (Chong). Using PCA analysis, exploratory data analysis can be undertaken with large datasets, including LCMS derived datasets, to determine relationships between sample groups. The addition of the shaded 95% confidence intervals to the PCA in **Figure 2.3A** indicates the similarity of FS1-FS4 relative to FS5. In this way, PCAs resolve general group relationships in large LCMS datasets. After general relationships are found, heatmaps can elucidate specific extracellular small molecules that differentiate groups (**Figure** 2.**3B**). Heatmaps can determine the concentration of an extracellular small molecule in a sample relative to the average concentration across the dataset. The color of a box at the intersection of a sample, usually in a column, and an extracellular small molecule, usually in a row, indicates the relative concentration. In **Figure 2.3B**, red boxes are extracellular small molecules with relatively high concentrations while blue shades indicate designate low concentrations. This analysis shows not only how samples are related, but what extracellular small molecules are important to discriminate groups. **Figure 2.3B** demonstrates this concept by highlighting the difference between FS5 and the other springs by indicating the extracellular small molecules that were present in higher concentrations in FS5 relative to FS1-FS4.

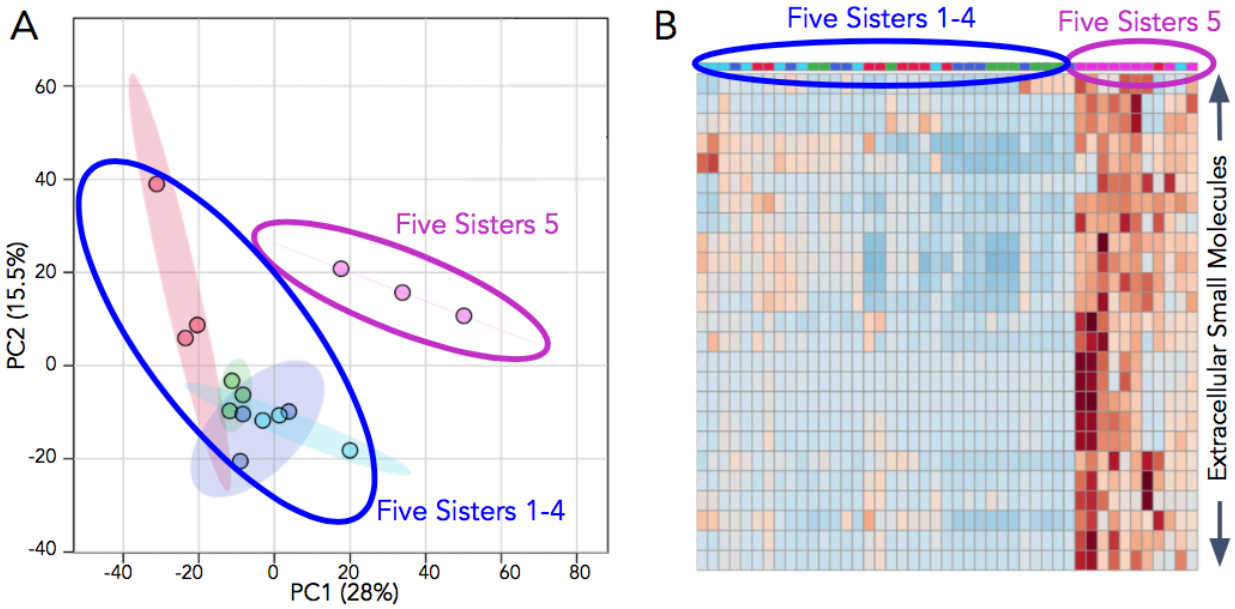


Figure 2.3. Extracellular small molecule data analysis techniques. (A) 2D-PCA scores plot of extracellular small molecules from Five Sisters 1-5. (B) Heatmap of the top 25 differentiating small molecules using the same data from A. Columns represent samples and rows represent extracellular small molecules. Red squares are relatively high concentration small molecules while blue squares are low concentration small molecules.

Extracellular small molecule data were further explored using many different analysis methods including the use of correlograms and Van Krevelen plots (Gaspar, Kim). Correlograms operate similar to heatmaps but allow for the integration of additional datasets using the R package mixOmixs (**Figure 2.4A**) (Rohart). In the analysis shown in **Figure 2.4A**, extracellular small molecule data are combined with 16s microbial data. The top 10 discriminating 16s microbial phylogenetic orders are paired with the top 10 discriminating extracellular small molecules. The box color of an intersection provides information as to the nature of the correlation between the specific 16s and extracellular small molecule variable. Blue boxes indicate positive correlations while red shades indicate negative correlations. In **Figure 2.4A**, it can be seen there is a strong negative correlation between Thermoproteales and the first selected extracellular small molecule. Per this analysis, as one of the variables, either Thermoproteales or the first extracellular small molecule, increases in concentration, the other one decreases. This correlation could be due to Thermoproteales using this small molecule metabolically and decreasing the environmental concentration along with many other scenarios. The box also has an asterisk which represents a significant correlation with a p-value <0.05. Van Krevelen plots are an analysis method used to group sets of dissolved organic material molecules into classes and determine the energy availability of aqueous systems (**Figure 2.4B**). Van Krevelen plots are a useful analysis tool that can reveal molecule classes and system energetics (Kim). Van Krevelen plots require formulas for the extracellular small molecules which are then used to calculate H:C and O:C ratios which can be used to determine classes of molecules in the aqueous system and energy availability (Smith). Assignment of formulas requires high accuracy and high-resolution MS typically found using a Fourier-transform ion cyclotron resonance type MS

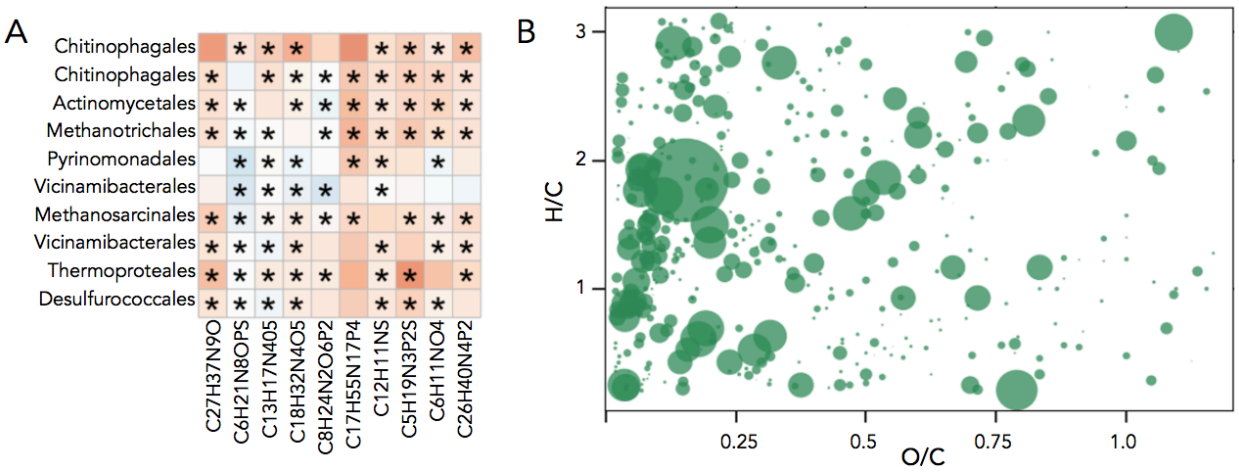


Figure 2.4. (A) Correlogram of the extracellular small molecules dataset and the 16s data from the same study. The top differentiating extracellular small molecules are compared to the top differentiating 16s ZOTUs to determine correlative values for each comparison. Red shaded squares indicate negative correlations while blue shaded squares indicate positive correlations. Stars indicate significant correlations to a p-value of <0.05. (B) Van Krevelen plot of the extracellular small molecules from Five Sisters 5. Points are based on the hydrogen/carbon and oxygen/carbon value of each molecular formula designation. On this Van Krevelen plot, point size indicates relative abundance. Different areas on the plot correlate to specific classes of macromolecules and plots can be compared to determine available energy sources for microbial populations.

(Gaspar). However, high-resolution LCMS based approaches have been successfully applied (Wang, Wang). After formula determination, H:C and O:C ratios were used as X and Y coordinates to plot each extracellular small molecule (Kim). This analysis can provide insight into the overall molecule class composition and allow for comparison between aqueous environments. With these data analysis methods, the environmental small molecule analysis provided a wealth of information on the thermal environment.

Intracellular small molecule, or metabolite, yields were greatly increased with optimized vortex and cell lysis procedures. Agitation most likely caused microbes residing on the sediment to become free in solution. Compared to samples analyzed on the LCMS without sonication or agitation with a vortex, an average of 894 more metabolites were found (**Table 2.1**). The optimal number of rounds of sonication and agitation was found to be two for each procedure. The increase in metabolites with the addition of sonication was stark, with two rounds of sonication increasing the number of metabolites by almost 700. While not as vital, one round of agitation increased the number of metabolites from 686 to 832. Further rounds of either sonication or agitation past two did not elicit a significant increase in intracellular small molecules as seen in **Table 2.1**.

After LCMS analysis, metabolomic profiles allowed for comparison between springs and years of sampling with 2-D PCA analysis (**Figure 2.5A**). Unlike the extracellular small molecule PCA analysis, the metabolite analysis does not show differences between the springs suggesting similar overall metabolic activity. Along with global changes, heatmaps and correlograms were able to elucidate specific metabolite variations and correlate selected metabolites to variables from other datasets similar to the extracellular small molecule analysis. In addition to the

Table 2.1. Unique intracellular molecules by treatment

Treatment	Features
Sonication-2 Vortex-0	686
Sonication-2 Vortex-1	832
Sonication-2 Vortex-2	894
Sonication-2 Vortex-3	866
Sonication-1 Vortex-2	867
Sonication-3 Vortex-2	787

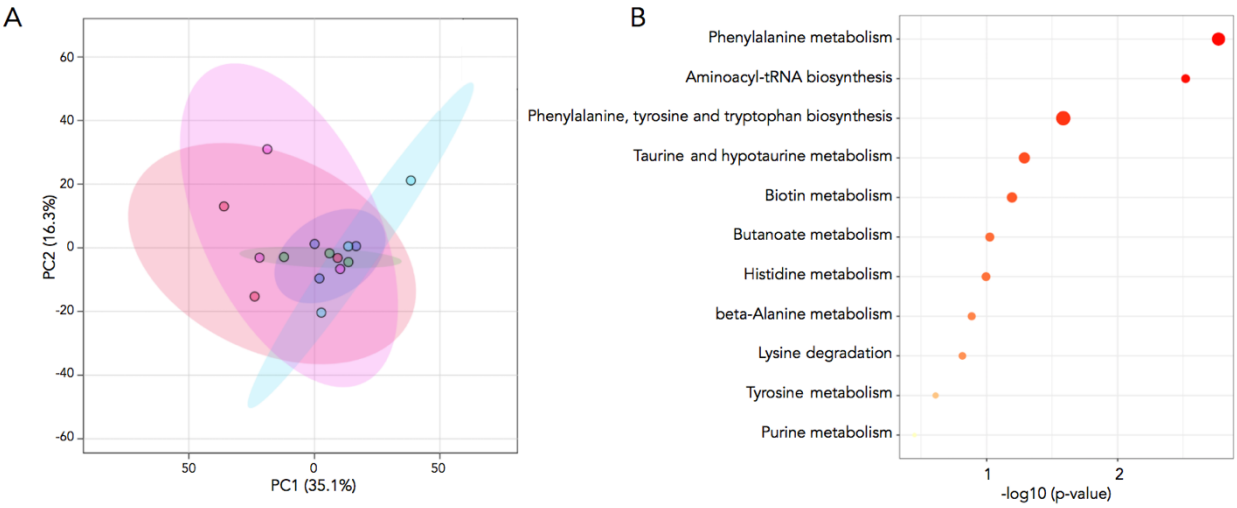


Figure 2.5. Intracellular small molecule data analysis. (A) PCA of intracellular small molecules. This is from the same data set that was analyzed in the extracellular small molecule analysis, yet unlike that dataset, Five Sisters 5 is not different with respect to intracellular small molecules. (B) Metabolic pathway analysis of the Five Sisters dataset showing pathways that differ between the springs. Dots represent differentially regulated metabolic pathways. The darker red the point, the more significant the p-value. The size of the dot also corresponds to the level of enrichment.

analysis techniques completed for the extracellular data; intracellular metabolic data allow for more detailed metabolic pathway analysis (**Figure 2.5B**). This type of analysis orders metabolites into metabolic pathways and determines the relative number of metabolites in a given pathway that exhibit significant variation between groups. This is a powerful tool that can indicate differences in metabolism between microbial populations over time or between springs.

Along with small molecule analysis, proteomic analysis of thermal sediment was optimized. Conventional protein digestions using trypsin were not effective at digesting thermophilic protein and little to no peptides were observed with LCMS analysis. A different protease, thermolysis, was selected and tested due to its thermostable nature. To generate an optimized thermolysin digestion protocol, digestion temperature, pH and time were tested on protein samples from Sulfolobus solfatericus, a thermophile common to acidic hot springs in YNP with temperatures greater than 75°C. Initial results confirmed that conventional protein digestion methods using trypsin were not effective at digesting thermophilic protein. A switch to the thermostable protease thermolysin allowed for higher digestion incubation temperatures and more complete digestion (**Figure 2.6A**). Three different temperatures, 65°C, 75°C and 85°C were tested for optimal digestion (**Figure 2.6B**). Although protein was digested, no significant difference was found with digestion in the temperature range of 65°C-85°C. pH was also tested for optimal digestion parameters. Digestions at pH 6, pH 7 and pH 8 were tested at 75°C and pH 6 clearly showed more efficient digestion with almost all protein digested after 30 minutes (**Figure 2.6C**). Given this analysis, we recommend a digestion temperature of 75°C at pH 6 for one hour (**Figures 2.6A-C**). Although most protein was digested after 30 minutes, increasing the time insured complete protein digestion. Using FASP protein preparation along with optimized

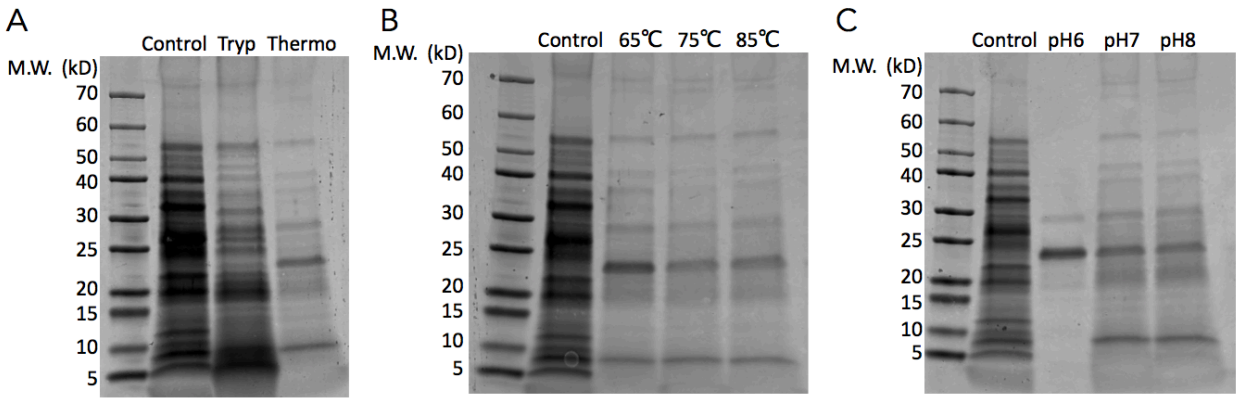


Figure 2.6. SDS-PAGE analysis of thermophilic protein digested by thermolysin under various digestion conditions. (A) 30-minute thermolysin digestion at pH 7 and 75℃ and standard trypsin digestion at 37℃ for 18 hours. (B) Temperature analysis testing digestions completed at 65℃, 75℃ and 85℃. All three digestions occurred at pH 7. (C) pH analysis testing digestions at pH 6, pH 7 and pH 8. Temperature was kept at 75℃ for these digestions.

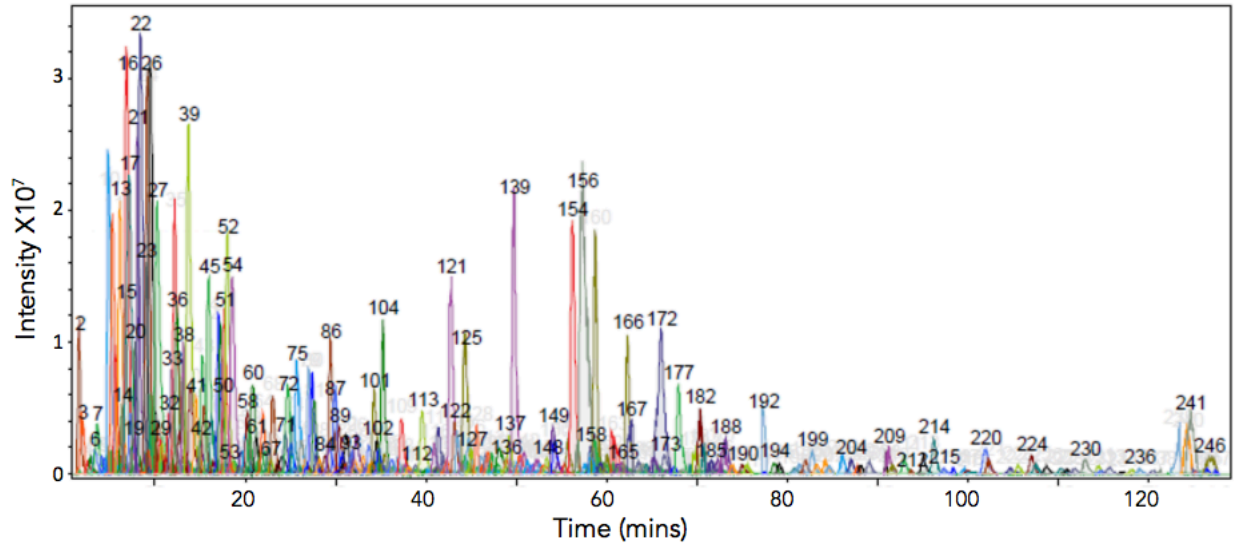


Figure 2.7. Proteomics analysis. Protein from a LCMS analysis of Octopus Spring in the White Creek Drainage. Proteomics analysis provided over 400 unique peptides and over 250 unique proteins per sample.

thermolysin digestion and a specific LCMS proteomics method, we were able to isolate over 400 unique peptide sequences from 15mL of thermal sediment (**Figure 2.7**).

Conclusion

Using the method developed herein, extracellular small molecule, intracellular small molecule and proteomic profiles for thermophilic microbes were generated more efficiently than conventional methods developed for mesophilic samples that are common in the literature. Sediment from hot springs can be challenging to obtain and the creation of comprehensive datasets from a small amount of sample can be challenging. However, with this method researchers can extract several different types of LCMS data to gain insight into thermal population ecology and thermal environments. A comprehensive dataset including intracellular and extracellular small molecule and proteomics data is invaluable to thermophile research. By optimizing several components pertinent to the analysis pipeline, a robust method was constructed which requires little starting material and provides a comprehensive view of thermophilic microorganisms and their environment.

Acknowledgements

The authors would like to thank the National Park Service for graciously allowing us to sample in YNP (permit YELL-2020-SCI-5480). We would also like to thank Jesse Thomas and Ganesh Balasubramanian of the Proteomics, Metabolomics and Mass Spectrometry Facility at Montana State University for their expertise. Funding for Proteomics, Metabolomics and Mass Spectrometry Facility used in this publication was made possible in part by the MJ Murdock Charitable Trust and the National Institute of General Medical Sciences of the National Institutes

of Health under Award Number P20GM103474. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would also like to thank the members of the Lawrence Lab and Dana Skorupa for providing cultured samples. Finally, we thank the WM Keck Foundation for providing funding for this research.

Literature Cited

BARNS, S. M., FUNDYGA, R. E., JEFFRIES, M. W. & PACE, N. R. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 1609-1613.

BARSNES, H. & VAUDEL, M. 2018. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *J Proteome Res,* 17**,** 2552-2555.

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. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology,* 30**,** 918-920.

CHONG, J., WISHART, D. S. & XIA, J. 2019. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics,* 68**,** e86.

DECASTRO, M. E., RODRIGUEZ-BELMONTE, E. & GONZALEZ-SISO, M. I. 2016. Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Frontiers in Microbiology,* 7**,** 21.

DITTMAR, T., KOCH, B., HERTKORN, N. & KATTNER, G. 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography-Methods,* 6**,** 230-235.

FARAJI, M., YAMINI, Y. & GHOLAMI, M. 2019. Recent Advances and Trends in Applications of Solid-Phase Extraction Techniques in Food and Environmental Analysis. *Chromatographia,* 82**,** 1207-1249.

GASPAR, A., HARIR, M., HERTKORN, N. & SCHMITT-KOPPLIN, P. 2010. Preparative free-flow electrophoretic offline ESI-Fourier transform ion cyclotron resonance/MS analysis of Suwannee River fulvic acid. *Electrophoresis,* 31**,** 2070-2079.

GROMIHA, M. M. & SURESH, M. X. 2008. Discrimination of mesophilic and thermophilic proteins using machine learning algorithms. *Proteins,* 70**,** 1274-9.

HEDLUND, B. P., MURUGAPIRAN, S. K., ALBA, T. W., LEVY, A., DODSWORTH, J. A., GOERTZ, G. B., IVANOVA, N. & WOYKE, T. 2015. Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. *Current Opinion in Microbiology,* 25**,** 136-145.

KEW, W., BLACKBURN, J. W. T., CLARKE, D. J. & UHRÍN, D. 2017. Interactive van Krevelen diagrams – Advanced visualisation of mass spectrometry data of complex mixtures. *Rapid Communications in Mass Spectrometry,* 31**,** 658-662.

KIM, S., KRAMER, R. W. & HATCHER, P. G. 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Analytical Chemistry,* 75**,** 5336-5344.

MUELLER, R. C., PEACH, J. T., SKORUPA, D. J., COPIÉ, V., BOTHNER, B. & PEYTON, B. M. 2020. An emerging view of the diversity, ecology, and function of Archaea in alkaline hydrothermal environments. *FEMS Microbiology Ecology*.

PLUSKAL, T., CASTILLO, S., VILLAR-BRIONES, A. & ORESIC, M. 2010. Mzmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics,* 11**,** 395.

ROHART, F., GAUTIER, B., SINGH, A. & LÊ CAO, K.-A. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PloS computational biology,* 13**,** e1005752-e1005752.

SANGUINETTI, M. & POSTERARO, B. 2016. Mass spectrometry applications in microbiology beyond microbe identification: progress and potential. *Expert Review of Proteomics,* 13**,** 965-977.

SMITH, H. J., TIGGES, M., D'ANDRILLI, J., PARKER, A., BOTHNER, B. & FOREMAN, C. M. 2018. Dynamic processing of DOM: Insight from exometabolomics, fluorescence spectroscopy, and mass spectrometry. *Limnology and Oceanography Letters,* 3**,** 225-235.

SPENCER, R. G. M., GUO, W., RAYMOND, P. A., DITTMAR, T., HOOD, E., FELLMAN, J. & STUBBINS, A. 2014. Source and biolability of ancient dissolved organic matter in glacier and lake ecosystems on the Tibetan Plateau. *Geochimica et Cosmochimica Acta,* 142**,** 64-74.

THE UNIPROT, C. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research,* 49**,** D480-D489.

VAN DER MEER, M. T. J., SCHOUTEN, S., BATESON, M. M., NÜBEL, U., WIELAND, A., KÜHL, M., DE LEEUW, J. W., SINNINGHE DAMSTÉ, J. S. & WARD, D. M. 2005. Diel Variations in Carbon Metabolism by Green Nonsulfur-Like Bacteria in Alkaline

Siliceous Hot Spring Microbial Mats from Yellowstone National Park. *Applied and Environmental Microbiology,* 71**,** 3978.

VAUDEL, M., BURKHART, J. M., ZAHEDI, R. P., OVELAND, E., BERVEN, F. S., SICKMANN, A., MARTENS, L. & BARSNES, H. 2015. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nature Biotechnology,* 33**,** 22-24.

WANG, K., ZHANG, Y., HUANG, R.-J., CAO, J. & HOFFMANN, T. 2018. UHPLC-Orbitrap mass spectrometric characterization of organic aerosol from a central European city (Mainz, Germany) and a Chinese megacity (Beijing). *Atmospheric Environment,* 189**,** 22-29.

WANG, X., HAYECK, N., BRÜGGEMANN, M., YAO, L., CHEN, H., ZHANG, C., EMMELIN, C., CHEN, J., GEORGE, C. & WANG, L. 2017. Chemical Characteristics of Organic Aerosols in Shanghai: A Study by Ultrahigh-Performance Liquid Chromatography Coupled With Orbitrap Mass Spectrometry. *Journal of Geophysical Research: Atmospheres,* 122**,** 11,703-11,722.

WINGEN, M., JAEGER, K. E., GENSCH, T. & DREPPER, T. 2017. Novel Thermostable Flavin-binding Fluorescent Proteins from Thermophilic Organisms. *Photochem Photobiol,* 93**,** 849-856.

CHAPTER THREE

LONGITUDINAL META-ANALYSIS OF THE FIVE SISTERS HOT SPRINGS IN YELLOWSTONE NATIONAL PARK REVEALS A DYNAMIC THERMOALKALINE ENVIRONMENT

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Jesse T. Peach

Contributions: Conceived the study, performed the extraction and MS analysis, interpreted the MS and meta-analysis data and wrote the manuscript.

Co-Author: Rebecca Mueller

Contributions: Collected samples, performed 16s analysis and assisted in manuscript preparation.

Co-Author: Dana Skorupa

Contributions: Collected samples, performed geochemical analysis and assisted in manuscript preparation.

Author: Margaux Mesle

Contributions: Collected samples and assisted in manuscript preparation.

Co-Author: Sutton Kanta

Contributions: Performed small molecule extractions.

Co-Author: Eric Boltinghouse

Contributions: Performed small molecule extractions.

Author: Bailey Sharon

Contributions: Performed small molecule extractions.

Co-Author: Valerie Copié

Contributions: Conceived the study and assisted in manuscript preparation.

Co-Author: Brian Bothner

Contributions: Conceived the study, interpreted the results and assisted in manuscript preparation.

Co-Author: Brent M. Peyton

Contributions: Collected samples, conceived the study, interpreted the results and assisted in manuscript preparation.

Manuscript Information

Jesse T. Peach, Rebecca Mueller, Dana Skorupa, Margaux Mesle, Sutton Kanta, Eric Boltinghouse, Bailey Sharon, Valerie Copié, Brian Bothner and Brent Peyton

Environmental Microbiology

Status of Manuscript:
X Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
_____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Wiley-Blackwell

Abstract

ABSTRACT-Thermoalkaline springs have been studied for the industrial possibilities offered by the high-pH, high temperature microbial populations that reside in them. However, the ecology of these springs is often overlooked. To better understand the functional outcomes of interactions between the geochemistry and the microbial community of these sites, we conducted a three-year study of the Five Sisters (FS) springs that included high resolution geochemical measures, metabarcoding of the bacterial and archaeal community with 16S rRNA sequencing, and mass spectrometry generated extracellular and intracellular small molecule data. By combining all four datasets, we completed a comprehensive meta-analysis of the intricate thermoalkaline spring system. Over the course of the study, microbial population dynamics responded to changing environmental conditions, with archaeal populations decreasing in both relative abundance and diversity when compared to bacterial populations. Decreases in Archaea relative abundance correlated with environmental changes leading to the decreased availability of specific nitrogen and sulfur containing extracellular small molecules. The multi-factorial analysis indicated a complex and dynamic environment which hosts an elastic microbial community able to adapt to changing geochemical and extracellular small molecule transitions with altered metabolism and an evolving population makeup.

Introduction

Thermoalkaline springs are unique environments of biological and potential industrial significance. Microbial populations harboring commercial applications have been well documented in these systems (Mueller). Current thermoalkaline enzyme research efforts are

broad and exemplified by the discovery of a suite of thermostable enzymes including lipolytic and hydrolytic enzymes (Lopez-Lopez, Sahay). Of particular interest is the development of thermostable cellulolytic enzymes capable of converting lignocellulosic biomass into sugars and ultimately ethanol under industrial conditions (Patel). The potential to develop thermo- and pH-stable enzymes for commercial applications is driving much of the research in this area (DeCastro, Verma), with much less focus on the fundamental ecology of these systems.

Although relatively rare, thermoalkaline springs can be found in many locations around the world, including Yellowstone National Park (YNP) where they are prevalent (Brock 1967, Christiansen 2001, Mueller 2020). One such group of springs in YNP includes the Five Sisters (FS) hot springs, located in the White Creek Drainage (WCD). WCD is part of the Lower Geyser Basin, the largest thermal basin in YNP. WCD, along with thermal features in the Norris Geyser Basin, have been identified as areas likely to contain distinct and dynamic environments (Rowe 1973, Fournier 1992).

As the number of described thermoalkaline springs has increased, new microbial lineages and metabolic networks have been uncovered (Lopez-Lopez 2015). Despite increased interest and investigation in these systems, gaps in knowledge remain. For example, the temporal dynamics of thermoalkaline microbial populations has not been investigated. In this study, the FS thermoalkaline springs were examined over the course of three years: 2017, 2018 and 2019. A meta-analysis was conducted using geochemical, metagenomic, and liquid chromatography mass spectrometry (LCMS) to determine temporal microbial trends and understand the driving factors behind shifts in population makeup and metabolism in this unique environment.

Materials and Methods

Description of the Five Sisters site

The FS site (RCN Database LWCG023A, LWCG023B, LWCG023C) consists of a group of alkaline-chloride springs located in the south-east corner of the Lower Geyser Basin in the WCD (44.5325°N, 110.7971°W) (**Figure 3.1A**). White Creek flows down the drainage and harbors several thermoalkaline springs including Spindle Geyser (LWCG149) and two of the best studied sites in YNP, Octopus Spring (LWCG138) and Mushroom Spring (Thiel). The FS site is located a few meters south of White Creek against a steeply inclined hill and consists of three springs with five distinct pools, labeled 1 to 5 from east to west, with variable interconnectivity (**Figure 3.1B**). FS1 is fed by a small geyser and is connected above-ground to FS2 and FS3. There is no visible above-ground connection between FS3 and FS4, although there may be below-ground connectivity between the two springs. FS4 and FS5 are connected above-ground and FS5 flow continues out of the spring group ultimately feeding into White Creek. FS1 is the largest spring at several meters wide and deep, followed by FS5, FS3, FS2 and FS4.

Sample collection

Samples were collected within one week of the first day of March in 2017, 2018 and 2019. Stainless steel cups on extendable poles (~0.25L total volume) were used to collect samples from each spring. The cups were rinsed with spring water first and then used to collect a small amount of sediment slurry. Samples were collected in triplicate from the same area of the spring for metabarcoding (16S rRNA gene sequencing) and mass spectrometry analysis. Sediment slurry samples were composed of ~8mL of sediment with ~7mL of thermal water and were placed in sterile 15mL conical centrifuge tubes after collection (Corning, Corning, NY). Slurry samples were immediately frozen in a dry ice and ethanol bath in the field and stored on

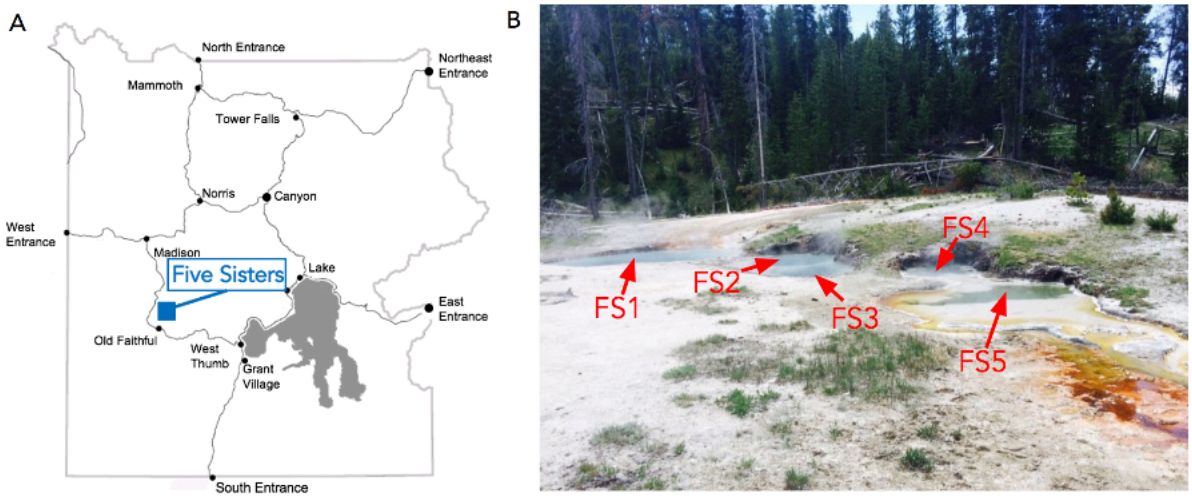


Figure 3.1. White Creek Drainage and Five Sisters location. (A) Map of YNP indicating the location of the Five Sisters springs north of Old Faithful. (B) Picture of the Five Sisters springs with all five pools labeled.

dry ice until transported to Montana State University (MSU) and placed in a -80°C freezer until genomic or mass spectrometry analysis.

Geochemical analysis

Aqueous geochemistry was monitored with each sampling event. Briefly, the temperature and pH of each site was measured *in situ* using a combined pH-temperature probe (Hach HQ30d, Hach Co., Loveland, CO). Dissolved oxygen was measured in the field using the High Range Dissolved Oxygen method and a portable colorimeter (Hach DR900, Hach Co., Loveland, CO). Total dissolved metals were analyzed using 0.22 μm filter-sterilized water acidified with 5% trace metal-grade nitric acid. Concentrations of total metals were quantified using an Agilent 7500ce ICP-MS by comparing to certified standards (Agilent Technologies, Environmental Calibration Standard 5183-4688) at MSU's Environmental and Biofilm Mass Spectrometry Facility. Samples for anion analysis were filtered through 0.22 μm filters, and the filtrate was analyzed using a Dionex ICS-1100 chromatography System (Dionex Corp., Sunnyvale, CA) equipped with a 25 μL injection loop and an AS22-4x250 mm anion exchange column, using an eluent concentration of 4.5 $mmol \cdot L^{-1}$ sodium carbonate and 1.4 $mmol \cdot L^{-1}$ sodium bicarbonate flowing at a rate of 1.2 $mL \cdot min^{-1}$. Samples for total carbon (TC), total nitrogen (TN), non-purgeable organic carbon (NPOC), and dissolved inorganic carbon (IC) were filtered through 0.22 μm filters, and the filtrate deposited in ashed glass vials filled with no headspace and capped with a septum. A Shimadzu TOC-CSH instrument with an attached TN module (Shimadzu Scientific Instruments, Columbia, MD) was used to measure TC/TN/NPOC/IC at MSU's Environmental Analytical Lab (EAL). Filtered samples acidified with sulfuric acid (final

pH<2) were also sent to the EAL for ammonium analysis using a Lachat QuickChem 8500 flow injection analyzer (Hach Co., Loveland, CO).

DNA extraction and metabarcoding analysis

DNA was extracted from each replicate using the FastDNA for Soil kit (MP Biomedicals) according to the manufacturer's instructions, with an additional 40s bead beating step. The V4 region of the 16S rRNA gene was targeted using the latest versions of the 515F-80R primers (Caparaso et al. 2011), 515F-A (GTGYCAGCMGCCGCGGTAA; Parada et al. 2016) and 806R-B (GGACTACNVGGGTWTCTAAT; Apprill et al. 2015) using Phusion Hot Start II Hi Fidelity polymerase in 25 ul reactions. Thermocycling conditions were an initial denaturation at 98^{o}C for 30 s, 22 cycles of 98^{o}C denaturation for 15 s, annealing at 58^{o}C for 30 s, extension at 72^{o}C for 20 s, with a final extension at 72^{o}C for 5 minutes. To facilitate multiplexing, dual-index barcodes were added in a second PCR using the Nextera kit (Illumina Inc.) with 10 cycles as above, but with an annealing temperature of 55^{o}C. PCR reactions were quantified using the Quant-It HS dsDNA kit (Invitrogen) and measured using a Biotek H2 plate reader. Reactions were pooled at equal concentrations and sequenced on a MiSeq using V3 600 cycle kits.

Paired-end reads were merged, primers were removed, and sequences were quality filtered with an expected error rate of 0.5 using USEARCH (Edgar 2010, Edgar 2016). Operational taxonomic units (OTUs) were identified using UNOISE3, which identifies biological sequences clustered into zero-radius OTUs (ZOTUs), similar to amplicon sequence variants. ZOTU tables were generated by mapping reads back to representative ZOTU

sequences. Taxonomy of ZOTUs was identified using the Ribosomal Database Project online classifier.

Sediment small molecule extraction

Sediment slurry samples were extracted resulting in intracellular and extracellular layers for LCMS analysis. Extraction began by thawing frozen samples in a 40ºC water bath and dislodging microbes from the sediment layer. Microbes were dislodged through two one-minute rounds of agitation on a vortex machine. Well-mixed samples were then centrifuged for five-minutes at 400 RPM to create a sediment pellet. The clear supernatant was then collected and placed in a clean vial and the procedure was repeated with the addition of a 5mL Milli-Q water wash. Wash supernatant was added to the original sediment-free supernatant. Combined supernatant was then centrifuged at 10,000 RPM for 15 minutes to create a cell pellet. The supernatant was collected in a clean vial for extracellular solid-phase extraction (SPE) analysis and enough phosphate buffer solution (1X PBS) was added to the cell pellet to completely cover the pellet.

The extracellular layer was first acidified to a pH of 2 using formic acid (Fisher Cemical, Hampton, ND) and Agilent Bond Elut PPL solid phase extraction cartridges (Agilent Technologies, Santa Clara, CA) were prepared. Preparation involved two cartridge volume additions of methanol (Fisher Chemical, Hampton, NH) followed by two cartridge volume additions of Milli-Q water and a final cartridge volume addition of methanol. SPE cartridges were then connected to a SPE manifold (VacMaster 10, Biotage, Uppsala, Sweden) and a vacuum pump to selectively concentrate and isolate extracellular small molecules. Acidified samples were passed through the cartridges and then placed under nitrogen to dryness. Clean

vials were placed below the cartridges and captured extracellular small molecules were eluted using 1mL of methanol. Extracellular small molecules were further concentrated under negative pressure using a Concentrator Plus (Eppendorf, Hamburg, Germany) until dry and then stored in a -80 ºC freezer until ready for LCMS analysis.

The vial containing the cell pellet was centrifuged at 400 RPM and the PBS supernatant was removed. Two cell pellet volumes of extraction buffer were added consisting of 8M urea (Fisher Chemical, Hampton, NH), 0.1M Tris-HCL (MilliporeSigma, Munich, Germany), 50mM ethylenediaminetetraacetic acid (EDTA) (MilliporeSigma, Munich, Germany) and 1X protease inhibitor mix (MilliporeSigma, Munich, Germany). Cell pellets were next lysed in a two-part procedure. First, cell pellets in extraction buffer were placed in liquid nitrogen for 10 seconds then removed and allowed to thaw. This procedure was repeated three times. After the freeze/thaw procedure, cell pellets underwent two rounds of sonication using a Biologics Inc. Ultrasonic Homogenizer 3000 (Bioloics, Manassas, VA) set at a 40% duty cycle for three minutes.

After cell lysis, the intracellular samples were centrifuged at 15,000 RPM for 30 minutes to pellet cell debris. The resulting supernatant was removed and placed in a clean vial while the debris were washed with the addition of 50µL of extraction buffer. After agitating the debris with a vortex machine, samples were centrifuged at 15,000 for 30 minutes. The second supernatant was again removed and added to the vial containing the first supernatant. Four volumes of ice-cold acetone (Fisher Chemical, Hampton, NH) was then added to precipitate protein. Samples were placed in a -80 ºC for two hours to aid precipitate formation. After two hours, samples were centrifuged at 5,000 RPM for five minutes and the intracellular supernatant

was collected and placed in a clean vial. As with the extracellular small molecule layer, the intracellular small molecule layer was concentrated under negative pressure to dryness and stored in a -80ºC freezer until LCMS analysis. When ready for LCMS analysis, both the intracellular and extracellular small molecule samples were reconstituted with 50µL of methanol:water (50:50) and placed in clean MS vials.

LCMS analysis

Samples were analyzed on an Agilent 6538 Q-TOF MS paired with an Agilent 1290 UHPLC (Agilent Technologies, Santa Clara, CA) using a 132Å, 2.2µm, 2.1mm X 150mm Cogent Diamond Hydride HPLC column (Microsolv, Greater Wilmington, NC) located in the Proteomics, Metabolomics and Mass Spectrometry facility at MSU. Ionization was accomplished via electrospray ionization in positive mode. Mobile phases A and B consisted of water with 0.1% formic acid and ACN with 0.1% formic acid, respectively. A 15-minute LC run time was used, starting with 100% of mobile phase B and moving to 30% B in a linear gradient over 14 minutes. At 14 minutes, mobile phase B was increased back to 100% for the final minute of the LC run. Flow was kept at 600µL/minute and the column compartment temperature was constant at 30°C.

LCMS raw data files were converted to .mzML files using MSConvert with Vendor peak picking (Chambers). Data mining was completed using mzMine (Pluskal). A minimum intensity of 1,000 counts was used throughout the mining process along with a ppm error of 20 and a time discrepancy of 0.1 minutes for determining unique peaks. Molecules formulas were determined using mzMine's formula generation identification feature with an error of 15ppm. After data mining, blank samples were used to remove residual features from the datasets. Features were

only kept in the experimental data if they had an area five times greater than the blank sample. Using this method, almost 5,000 combined features were found in the extracellular small molecule samples and over 4,500 features were found in the combined intracellular small molecule samples. Datasets were then grouped and analyzed using MetaboAnalyst (Chong). Samples were removed if not in over 50% of the samples and the data were filtered using their IQR value and normalized using the autoscale function.

Correlation analysis

The geochemical, 16S, intracellular and extracellular small molecule datasets were compared using correlograms. This was accomplished by utilizing Bioconductor and mixOmics packages in R (Huber, Rohart). Two datasets were compared at a time and the top 10 discriminating features were determined using a partial least-squares discriminating analysis (PLSDA). The top 10 features from each dataset were then compared between springs and years to determine their correlative relationship. An ANOVA was also conducted for each relationship to determine the significance of the correlation.

Results

Temporal geochemistry of the FS site

An analysis of the geochemical data demonstrated similarities and differences between springs and years (**Table 3.1**). A principal component analysis (PCA), 2D-PCA scores plot with shaded 95% confidence intervals (**Figure 3.2A),** revealed similar geochemical environments between 2017, 2018 and 2019. Annual geochemical fluctuations appeared to be moderate although an analysis using hierarchical clustering grouped each set of springs by year with the

Table 3.1. Geochemical characteristics from 2017-2019

Year	2017					2018					2019				
Site	FS1	FS2	FS3	FS4	FS5	FS1	FS2	FS3	FS4	FS5	FS1	FS2	FS3	FS4	FS5
SWE			88%					110%					124%		
Temp (C)	78	78	75	77	66	71	74	73	71	64	75.1	80.3	79.8	80	70
pH	9	8	8	8	8	9	8.5	8.5	8.5	8.5	8.8	8.4	8.4	8.4	8.5
DO (mg/ L)	1.7	1.3	1.2	1.3	1.5	0.2	0.2	0.7	0.2	0.8	2.2	0.4	0.5	0.5	4.4
Chloride (mg/L)	216	224.6	234.3	243.1	246.9	253.4	256.2	263.4	255.1	256.5	277.1	296.1	277.2	272	291.1
Sulfate (mg/L)	12.4	12.8	13.3	13.6	14	14.3	14.6	14.7	14.8	15.2	11.7	11.3	11.5	11.4	12
Sodium (mg/L)	158.8	143.3	144.2	157.4	157	140.8	152.4	141.7	161.2	158.1	169.3	170.1	173.6	170	168.2
Potassium (mg/L)	6.9	6.6	6.9	6.7	6.7	6.4	6.8	6.2	7.1	7.2	6.3	6	7.2	7.2	7.4
Zinc (ug/L)	9.5	12.1	11.4	11.5	12	22.4	18.4	19.2	21.9	14.6	19	22	20	16	18
Arsenic (ug/L)	750	670	677	731	715	698	740	667	771	771	800	790	770	780	770
Molybdenum (ug/L)	11.7	11.8	12.3	11.8	11.8	11.7	12.4	11.2	13	13	12	13	12	12	13
TC (mg/L)	45.3	55.3	55.2	55.3	55.7	44.7	54.5	54.1	54.8	54.4	45.7	57	56.6	56	55.4
TN (ppm)	0.20	0.08	0.08	0.09	0.09	0.05	0.06	0.06	0.05	0.05	0.13	0.1	0.1	0.1	0.1
Inorganic Carbon (mg/L)	43.9	54.4	55.2	54.4	55.5	45.3	55	52.4	52.9	54.9	44.1	53.6	53.7	54.3	53.7
NPOC (mg/L)	1.1	0.3	0.2	0.4	0.3	0.4	0.5	0.3	0.5	0.3	0.8	0.6	0.6	0.6	0.5
NH_4 (ug N/L)	25	23.1	21.5	23.6	27.8	14	29	20	24	17	32	26	26	24	17

SWE denotes snow water equivalent percentage of the median for the previous year. DO is dissolved oxygen. TC is total carbon. TN is total nitrogen. NPOC is non-purgeable organic carbon.

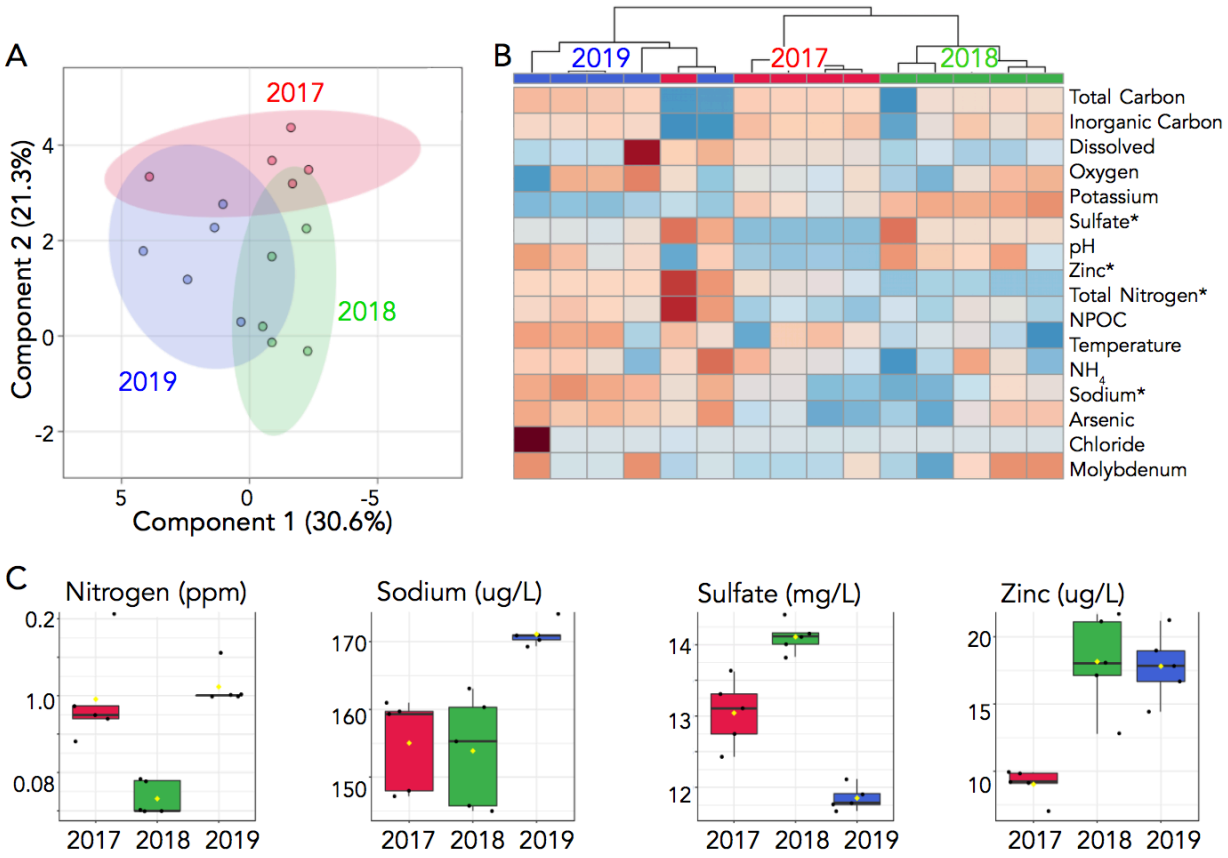


Figure 3.2. Geochemistry analysis. (A) 2D-PCA scores plot of all five springs by year. (B) Heatmap visualization of geochemical features for all 5 springs grouped by years as shown in columns. Asterisks indicate statistically significant geochemical features accounting for the three different groups (C) One-way parametric ANOVA analysis indication geochemical features whose levels are significantly different (p-value <0.05) between the three years.

exception of FS in 2017 and 2018 (**Figure 3.2B**). Four geochemical measurements, nitrogen, sodium, sulfate and zinc concentrations, were significantly different between the three years with ANOVA p-values <0.05 (**Figure 3.2B, 3.2C**). However, the patterns of concentration changes between 2017 and 2019 for these variables was not linear. Total nitrogen concentrations decreased in 2018 and then increased back to 2017 levels in 2019 (**Figure 3.2C**). Sodium and zinc both increased in 2019 relative to 2017 but stayed constant between 2017 and 2018 for sodium and 2018 and 2019 for zinc. The pattern of sulfate was also different, increasing in 2018 and then decreasing in 2019 to below 2017 levels. Table 1 also indicates that snowpack was variable between years. Snow water equivalents (SWE) in 2016 were slightly lower than average while 2017 and 2018 were higher than normal with 2018 having higher SWE than 2017.

16s rRNA profiles

16s rRNA ZOTU relative abundance was analyzed using MicrobiomeAnalyst to determine year and spring specific microbial trends (Chong). This analysis revealed several microbial patterns including two striking population trends (**Figure 3.3A**). First, a consistent decrease was observed in the relative abundance of microbial ZOTUs from phylum Thermoproteota from 2017 to 2018 and from 2018 to 2019. The second was an increase in relative abundance of ZOTUs from the bacterial phyla Proteobacteria and Deinococcota. Over the course of the study, FS1 and FS5 had a marked increase in Deinococcota relative abundance while FS2, FS3 and FS4 had relative Proteobacteria population increases in 2019 (**Figure 3.3B**). Other trends, such as a relative decrease in Acidobacteriota populations in FS3, FS4 and FS5 from 2017 to 2019, were also apparent. Along with a decrease in abundance, archaeal alpha

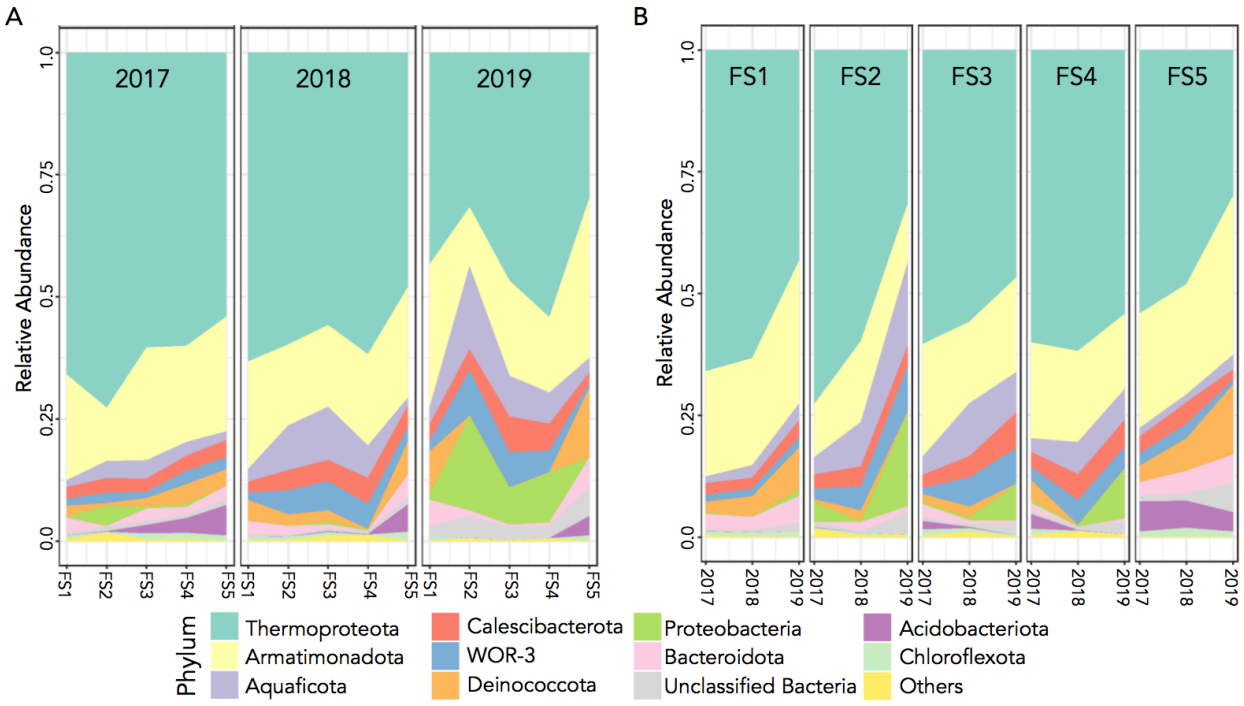


Figure 3.3. Relative microbial abundance. (A) Phylum-level relative abundance of each spring indicating the microbial population trends for each specific spring over the three-year period. Each panel represents a year from 2017-2019 with springs FS1-FS5 relative abundance levels. (B) Phylum-level abundance for each year in the study. Each panel represents a specific spring over the three-year period.

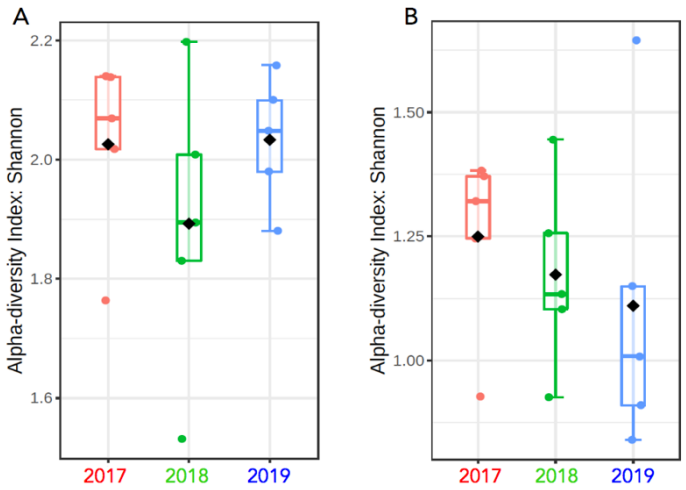


Figure 3.4. Microbial alpha diversity using a Shannon Index for all springs in a given year. (A) Bacterial alpha diversity from 2017-2019. (B) Archaeal alpha diversity from 2017-2019. Shannon Index analysis accounts for both the evenness and richness in a system.

diversity also decreased in each spring between 2017 and 2019 (**Figure 3.4**). The exception to this was FS5, which showed an increase in alpha diversity in 2019.

Small molecule profiles

As the geochemical results did not yield likely factors explaining the archaeal loss seen in the 16s data, our next step was to profile the intracellular small molecules. This analysis of intracellular small molecules extracted from cell pellets suggested a metabolic transition had occurred. A 2D-PCA score plot of the global metabolome for each year indicated a contrast in metabolic activity between 2017 and both 2018 and 2019 (**Figure 3.5A**). Intracellular small molecules from 2018 and 2019 were somewhat different but were more closely related than those from 2017. Specific identified features that most discriminate the groups using an ANOVA include hypotaurine and butyric acid (**Supplemental Figure 3.1A**). Supplemental Figure 1A also indicates that springs cluster by year using the top 10 discriminating intracellular features.

Extracellular small molecules from the sediment exhibited a similar trend wherein 2017 is markedly different than 2018 and 2019 (**Figure 3.5B**). The variation between years was most pronounced in the 2017 FS1 and FS5 sediment environments. Exploring the small molecules that most differentiate the springs and years via a heatmap indicated the abundance of specific sulfur and nitrogen containing species in spring sediment from 2017 (**Supplemental Figure 3.1B**). There was a general trend indicating a decrease in the number of nitrogen, sulfur, and combined nitrogen and sulfur containing compounds in the thermal sediment each year in the study (**Figure 3.5C**). Sulfur containing compounds, in particular, had a 30% decrease between 2017 and 2019.

Statistical correlations

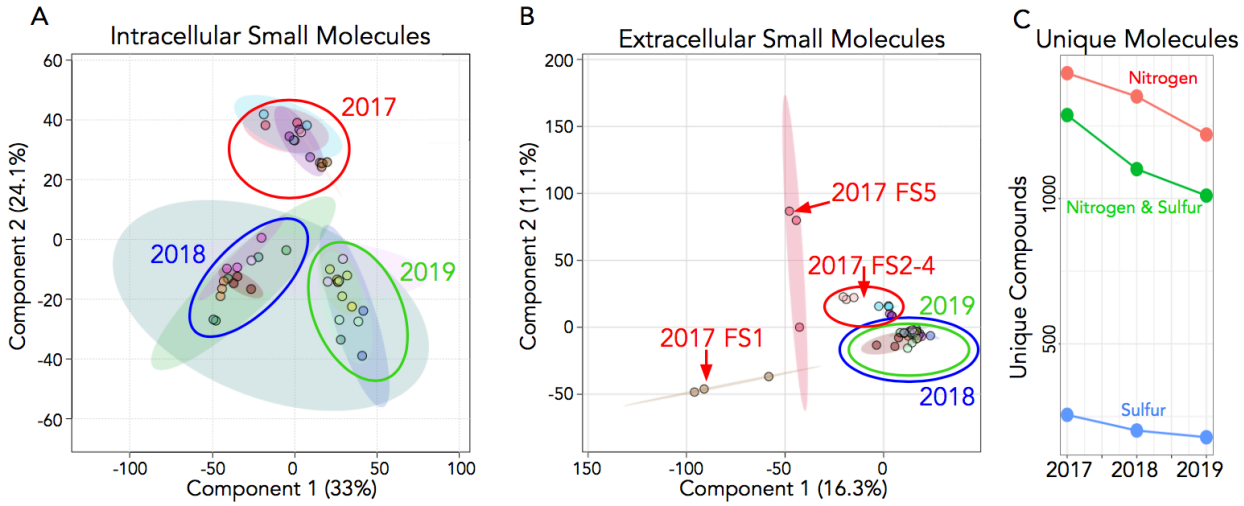


Figure 3.5. Intracellular and extracellular small molecules analysis. (A) PCA of the intracellular small molecules. Metabolites from 2017 group separately from 2018 and 2019. Metabolites from 2018 and 2019 group independently but overlap in the 95% confidence intervals for several springs, most strongly in FS4 and FS5. (B) PCA for extracellular small molecules in the sediment. 2018 and 2019 group very closely together relative to 2017. FS2-4 from 2017 group together and are closer to 2018 and 2019 than FS1 and FS5 from 2017. (C) Plot showing unique sulfur, nitrogen and sulfur and nitrogen containing molecules. Element specific molecules are shown in each year from 2017 to 2019. Formulas were determined from the mass spectrometry datasets using an error of 15ppm.

To determine if there were correlations between statistically important variables in the four datasets, correlograms were generated. Correlograms identify variables that discriminate groups, in this case by year, and then calculate the strength and statistical significance of the correlation. The top 10 most discriminating variables between year and springs were selected from the small molecule datasets and correlated with 16s rRNA data (**Supplemental Figures 3.2A-C**). With respect to the correlogram with geochemistry, seven of 10 ZOTUs exhibited a significant positive correlation with sulfate concentrations (**Supplemental Figure 3.2A)**. Zinc concentrations were similarly associated with an increase in the top 10 microbes, though this was not as ubiquitous as sulfate. Sodium, arsenic and dissolved oxygen showed an opposite trend having a negative correlation with most of the selected ZOTUs. Small molecule correlograms indicated the strong correlation between specific sulfur and nitrogen containing compounds and Archaea. Both intra- and extracellular correlograms with the ZOTU data were dominated by nitrogen and sulfur compounds (**Figure 3.6, Supplemental Figures 3.2B-C)**.

A specific examination was conducted using ZOTU data from archaea and only datasets with nitrogen and sulfur containing small molecules to determine what could be leading to the relative loss of Archaea. This analysis showed strong positive correlations between specific nitrogen and sulfur containing compounds and Archaea (**Figure 3.6**). Not only were almost all correlations positive, i.e. increases in specific small molecules were observed with increases in archaeal abundance, but the majority of the correlations were significant with a p-value <0.05.

Discussion

This study was undertaken to explore thermoalkaline environments and the microbial life that inhabits them. The described meta-analysis allowed for the most comprehensive view of

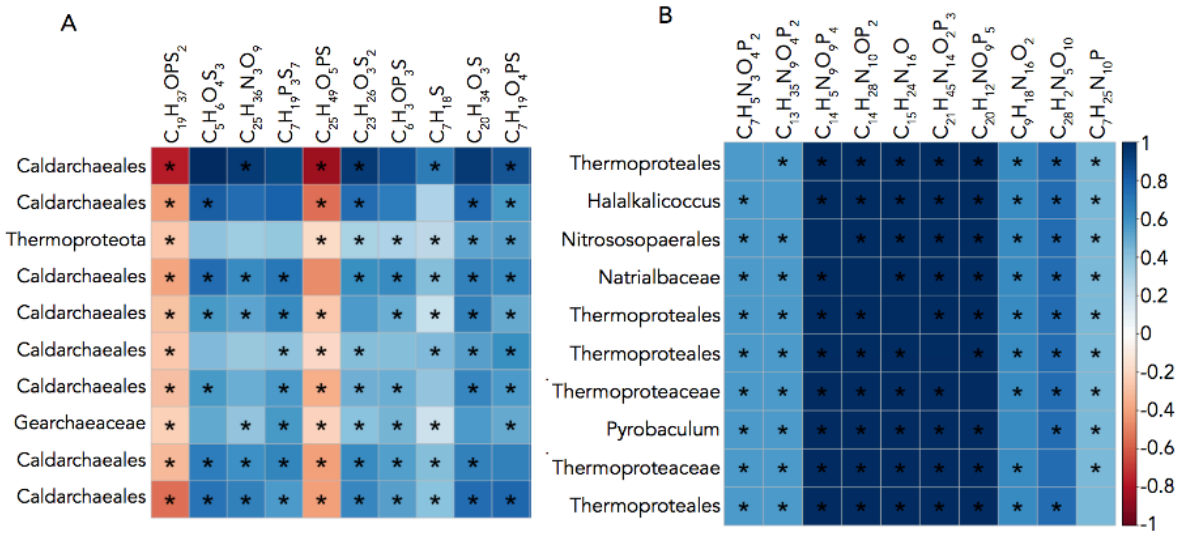


Figure 3.6. Correlation analysis. (A) Correlogram as described in the methods section investigating archaeal ZOTUs and sulfur containing extracellular small molecules. (B) Correlogram of archaeal ZOTUs and nitrogen containing extracellular small molecules. The intensity of the color indicates the strength of the correlation, with blue demonstrating positive and red demonstrating negative correlations. An asterisk indicates a p-value of <0.05 for the correlation.

microbial life in the FS spring system. Our data indicate that a system change likely occurred between 2017 and 2018 which significantly impacted the environment and microbial ecology of the FS springs. Samples from 2017 to 2018 displayed a coordinated extracellular small molecule, microbial and intracellular small molecule shift.

Microbial changes across the term of the study were defined by decreases in relative abundance of archaeal microbes in 2018 and then again in 2019, prominently indicated by a decline in relative abundance of Thermoproteota. Abundance decreases were also associated with reduced alpha diversity. This is in contrast with the bacterial populations which increased in overall relative abundance in 2018 and 2019 and remained at a somewhat consistent alpha diversity level. This phenomenon has been observed by Pala, et al., where different archaeal and bacterial abundance shifts were the result of differing environmental factors in the same aqueous environment.

Microbial metabolic changes were correlated to a combination of environmental factors. However, the geochemical variables with significant differences over the course of the study did not exhibit a characteristic a pattern that reflected the microbial data. Our analysis provided evidence that the strongest correlations belong to associations with sulfur and nitrogen containing small molecules. These specific small molecules decreased steadily from 2017 to 2019 and coincided with the general decrease in the relative abundance and diversity of Archaea. Intracellular small molecule profiles also shifted globally between 2017 and 2018, suggesting a metabolic change of the spring system during this time period.

Archaeal metabolism, along with bacterial metabolism, is known to be important in the nitrogen and sulfur cycles of environments, especially in extreme environments (Offre). Our

analysis implies that a specificity to the nitrogen and sulfur containing metabolically active compound requirements of Archaea relative to Bacteria may also exist. These observations lead to the hypothesis that environmental shifts led to different small molecules with contrasting elemental composition and structure being available, which required thermophiles in the FS system to adapt their metabolic strategies. Alterations in optimal metabolism then led to differential success of specific organisms in the FS system and a shift in population makeup occurred from 2017 to 2019 as defined by a decline in the relative abundance of Archaea. Although the decline in nitrogen and sulfur containing extracellular small molecules could be due to a microbial influence, Gonisor, et al., concluded that the chemodiversity seen in Octopus Spring was likely not microbially driven. Their conclusion was based on the low biodiversity in thermoalkaline springs and the high level of unique extracellular small molecules.

One potential explanation is the differential annual mixing of ground and surface water in the springs, where runoff, groundwater and snowpack can impact spring geochemistry and extracellular small molecule composition (Gibson, Gonsior). In our study, samples were taken in the winter with direct runoff long over, such that the environment should be fairly stable at the time of collection. However, the amount of groundwater recharge from the previous year may have a significant impact on resulting spring water composition the following winter. The area around WCD had unusually high snowpack in 2017 and 2018 (impacting 2018 and 2019 data) relative to the previous and following years according to the Water Resources Data System & State Climate Office of Wyoming (www.wrds.uwyo.edu). On March 1st, 2017 and 2018, the snow water equivalent was 110% and 124% of the median average for the time of year in, respectively. Snow water equivalent in 2016 was 88% of the median, potentially impacting 2017

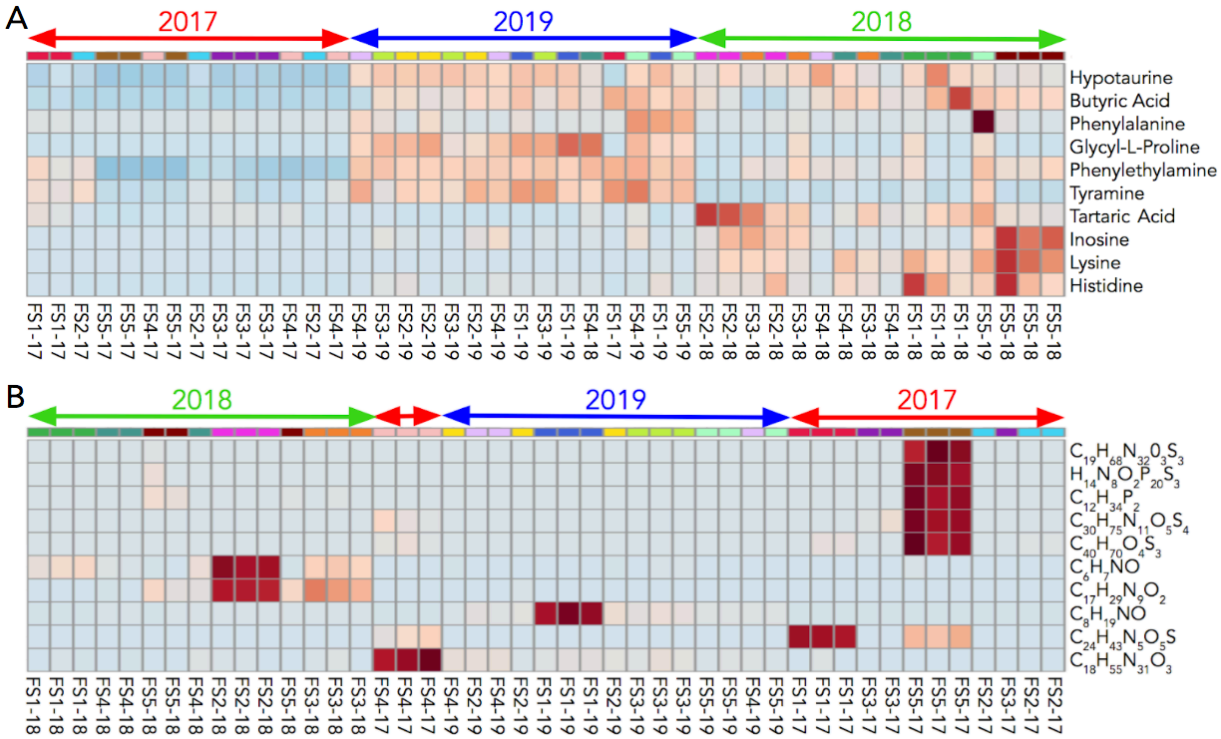
data. This trend indicating an increased snowpack from 2016 to 2018 correlates with the loss of unique nitrogen and sulfur containing compounds and archaeal decline. During the increased runoffs in the 2017 and 2018 season from high snowpack, the spring composition likely went through a significant change leading to distinct differences in the environmental small molecule structure.

Our analysis characterized the complexity and dynamism of thermoalkaline spring ecology. It also revealed how thermophiles, in this case Archaea, can survive in such extreme conditions yet may be sensitive to small environmental perturbations. The combination of traditional geochemical techniques combined with mass spectrometry analysis exposed the limitations of only using elemental composition and traditional analyses to describe complex environmental systems. Mass spectrometry analysis indicated specific compound changes that correlate with archaeal relative abundance in thermoalkaline systems that would have been missed using a purely geochemical examination. Our comprehensive analysis revealed that an environmental event occurred between 2017 and 2019 that shifted spring compound structure, manifesting in sulfur and nitrogen containing compound transitions, leading to metabolic changes and a relative decrease in Archaea.

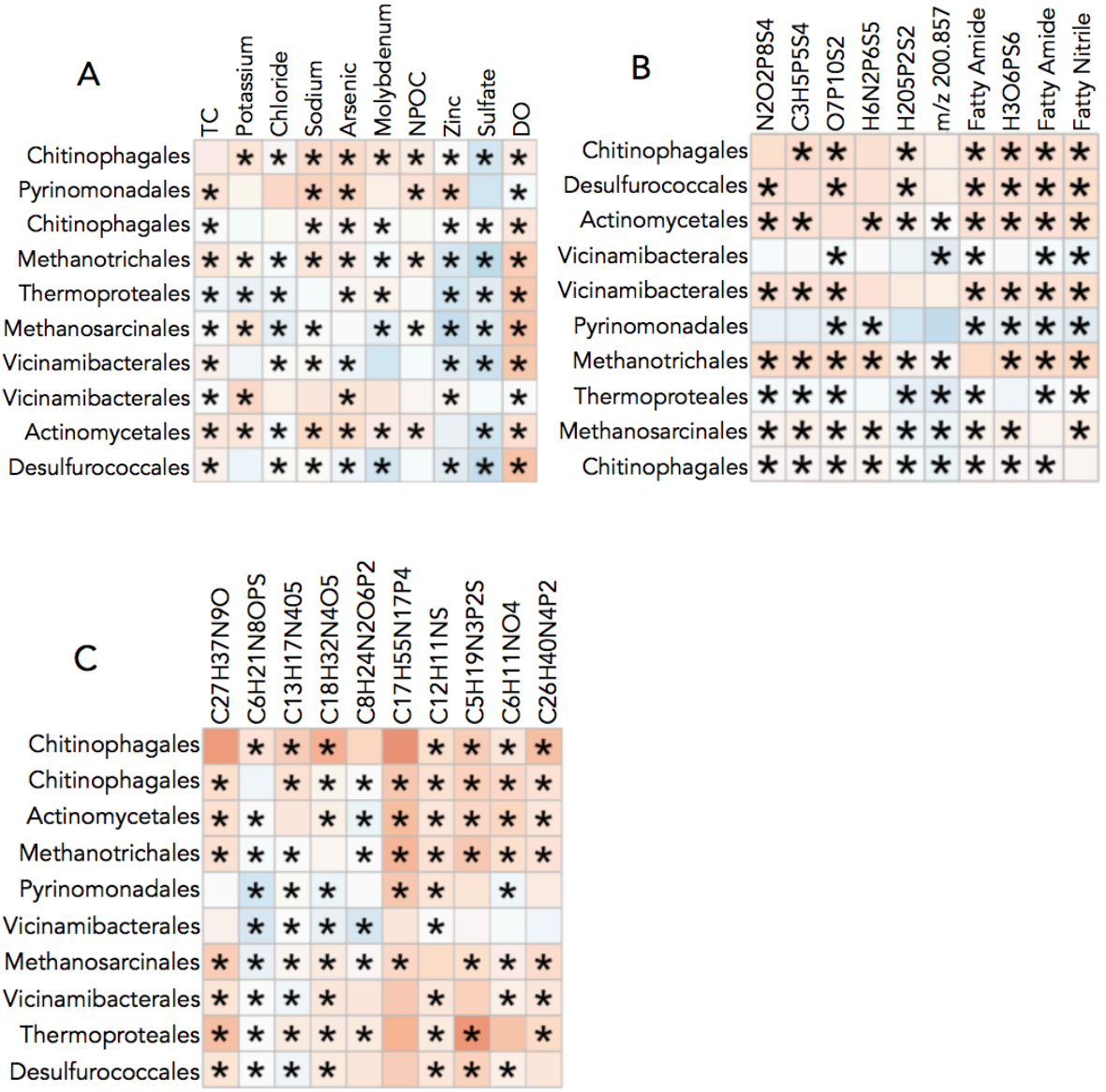
Acknowledgements

The authors would like to thank the National Park Service for graciously allowing us to sample in YNP (permit YELL-2017-SCI-5480 to YELL-2020-SCI-5480) We would also like to thank Jesse Thomas and Ganesh Balasubramanian of the Proteomics, Metabolomics and Mass Spectrometry Facility at Montana State University for their expertise. Funding for Proteomics, Metabolomics and Mass Spectrometry Facility used in this publication was made possible in part

by the MJ Murdock Charitable Trust and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM103474. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Finally, we thank the WM Keck Foundation for providing funding for this research.



Supplemental Figure 3.1. Small molecule heatmaps. (A) Heatmap showing the top 10 most discriminating intracellular small molecules. Columns indicate a specific spring in a specific year and rows represent small molecules. A red square indicates a high concentration relative to the average while blue squares indicate a low relative concentration. The darker squares indicate higher concentrations. (B) Heatmap showing the top 10 most discriminating extracellular small molecules.



Supplemental Figure 3.2. Correlograms. (A) Correlogram of the 16s data and the geochemical data. (B) Metagenomic data paired with intracellular small molecules. (C) Metagenomic data and extracellular small molecules data. Blue squares indicate positive correlations while red squares indicate negative correlations. An asterisk is shown if the correlation has a p-value of <0.05.

Literature Cited

APPRILL, A., MCNALLY, S., PARSONS, R. & WEBER, L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology,* 75**,** 129-137.

BROCK, T. D. 1967. Life at High Temperatures. *Science,* 158**,** 1012.

CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., LOZUPONE, C. A., TURNBAUGH, P. J., FIERER, N. & KNIGHT, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 4516-4522.

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. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology,* 30**,** 918-920.

CHONG, J., LIU, P., ZHOU, G. & XIA, J. 2020. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols,* 15**,** 799-821.

CHONG, J., WISHART, D. S. & XIA, J. 2019. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics,* 68**,** e86.

CHRISTIANSEN, R. L. 2001. The Quaternary and Pliocene Yellowstone Plateau volcanic field of Wyoming, Idaho, and Montana. *Professional Paper.* - ed.

DECASTRO, M.-E., RODRÍGUEZ-BELMONTE, E. & GONZÁLEZ-SISO, M.-I. 2016. Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Frontiers in microbiology,* 7**,** 1521-1521.

EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* 26**,** 2460-2461.

EDGAR, R. C. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv,* Preprint.

FOURNIER, R., THOMPSON, MICHAEL J., HUTCHINSON, RODERICK A. 1992. The geochemistry of hot spring waters at Norris Geyser Basin, Yellowstone National Park. *International symposium on water-rock interactions.*

GIBSON, M. L. & HINMAN, N. W. 2013. Mixing of hydrothermal water and groundwater near hot springs, Yellowstone National Park (USA): hydrology and geochemistry. *Hydrogeology Journal,* 21**,** 919-933.

GONSIOR, M., HERTKORN, N., HINMAN, N., DVORSKI, S. E. M., HARIR, M., COOPER, W. J. & SCHMITT-KOPPLIN, P. 2018. Yellowstone Hot Springs are Organic Chemodiversity Hot Spots. *Scientific Reports,* 8**,** 14155.

HUBER, W., CAREY, V. J., GENTLEMAN, R., ANDERS, S., CARLSON, M., CARVALHO, B. S., BRAVO, H. C., DAVIS, S., GATTO, L., GIRKE, T., GOTTARDO, R., HAHNE, F., HANSEN, K. D., IRIZARRY, R. A., LAWRENCE, M., LOVE, M. I., MACDONALD, J., OBENCHAIN, V., OLEŚ, A. K., PAGÈS, H., REYES, A., SHANNON, P., SMYTH, G. K., TENENBAUM, D., WALDRON, L. & MORGAN, M. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods,* 12**,** 115-121.

LÓPEZ-LÓPEZ, O., CERDÁN, M.-E. & GONZÁLEZ-SISO, M.-I. 2015. Thermus thermophilus as a Source of Thermostable Lipolytic Enzymes. *Microorganisms,* 3**,** 792-808.

MUELLER, R. C., PEACH, J. T., SKORUPA, D. J., COPIÉ, V., BOTHNER, B. & PEYTON, B. M. 2020. An emerging view of the diversity, ecology, and function of Archaea in alkaline hydrothermal environments. *FEMS Microbiology Ecology*.

OFFRE, P., SPANG, A. & SCHLEPER, C. 2013. Archaea in biogeochemical cycles. *Annu Rev Microbiol,* 67**,** 437-57.

PALA, C., MOLARI, M., NIZZOLI, D., BARTOLI, M., VIAROLI, P. & MANINI, E. 2018. Environmental Drivers Controlling Bacterial and Archaeal Abundance in the Sediments of a Mediterranean Lagoon Ecosystem. *Current Microbiology,* 75**,** 1147-1155.

PARADA, A. E., NEEDHAM, D. M. & FUHRMAN, J. A. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology,* 18**,** 1403-1414.

PATEL, A. K., SINGHANIA, R. R., SIM, S. J. & PANDEY, A. 2019. Thermostable cellulases: Current status and perspectives. *Bioresour Technol,* 279**,** 385-392.

PLUSKAL, T., CASTILLO, S., VILLAR-BRIONES, A. & ORESIC, M. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics,* 11**,** 395.

ROHART, F., GAUTIER, B., SINGH, A. & LÊ CAO, K.-A. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology,* 13**,** e1005752-e1005752.

ROWE, J. J., FOURNIER, R. & MOREY, G. Chemical analysis of thermal waters in Yellowstone National Park, Wyoming, 1960-65. 1973.

SAHAY, H., YADAV, A. N., SINGH, A. K., SINGH, S., KAUSHIK, R. & SAXENA, A. K. 2017. Hot springs of Indian Himalayas: potential sources of microbial diversity and thermostable hydrolytic enzymes. *3 Biotech,* 7**,** 118.

THIEL, V., WOOD, J. M., OLSEN, M. T., TANK, M., KLATT, C. G., WARD, D. M. & BRYANT, D. A. 2016. The Dark Side of the Mushroom Spring Microbial Mat: Life in the Shadow of Chlorophototrophs. I. Microbial Diversity Based on 16S rRNA Gene Amplicons and Metagenomic Sequencing. *Frontiers in Microbiology,* 7.

VERMA, P., AJAR NATH YADAV, LIVLEEN SHUKLA, ANIL KUMAR SAXENA AND ARCHNA SUMAN 2015. Hydrolytic enzymes production by thermotolerant Bacillus altitudinis IARI-MB-9 and Gulbenkiania mobilis IARI-MB-18 isolated from Manikaran hot springs. *International Journal of Advanced Research,* 3**,** 1241-1250.

CHAPTER FOUR

TEMPORAL METABOLIC RESPONSE YIELDS A DYNAMIC BIOSIGNATURE OF INFLAMMATION

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Jesse T. Peach

Contributions: Conceived the study, performed MS analysis, interpreted MS data, completed statistical analysis and wrote the manuscript.

Co-Author: Stephanie M. Wilson

Contributions: Collected blood samples, performed cytokine analysis and assisted in manuscript preparation.

Co-Author: Logan D. Gunderson

Contributions: Performed serum extractions and assisted in MS analysis.

Author: Lizzi Frothingham

Contributions: Performed serum extractions and assisted in MS analysis.

Co-Author: Tan Tran

Contributions: Assisted in statistical results interpretation.

Co-Author: Seth T. Walk

Contributions: Conceived the study.

Author: Carl J. Yeoman

Contributions: Conceived the study and assisted in manuscript preparation.

Co-Author: Brian Bothner

Contributions: Conceived the study, interpreted the results and assisted in manuscript preparation.

Co-Author: Mary P. Miles

Contributions: Conceived the study, collected blood samples, interpreted the results and assisted in manuscript preparation.

Manuscript Information

Jesse T. Peach, Stephanie M. Wilson, Logan D. Gunderson, Lizzi Frothingham, Tan Tran, Seth T. Walk, Carl J. Yeoman, Brian Bothner, Mary P. Miles

iScience

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
__X__ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Cell
Submitted 01/19/2021

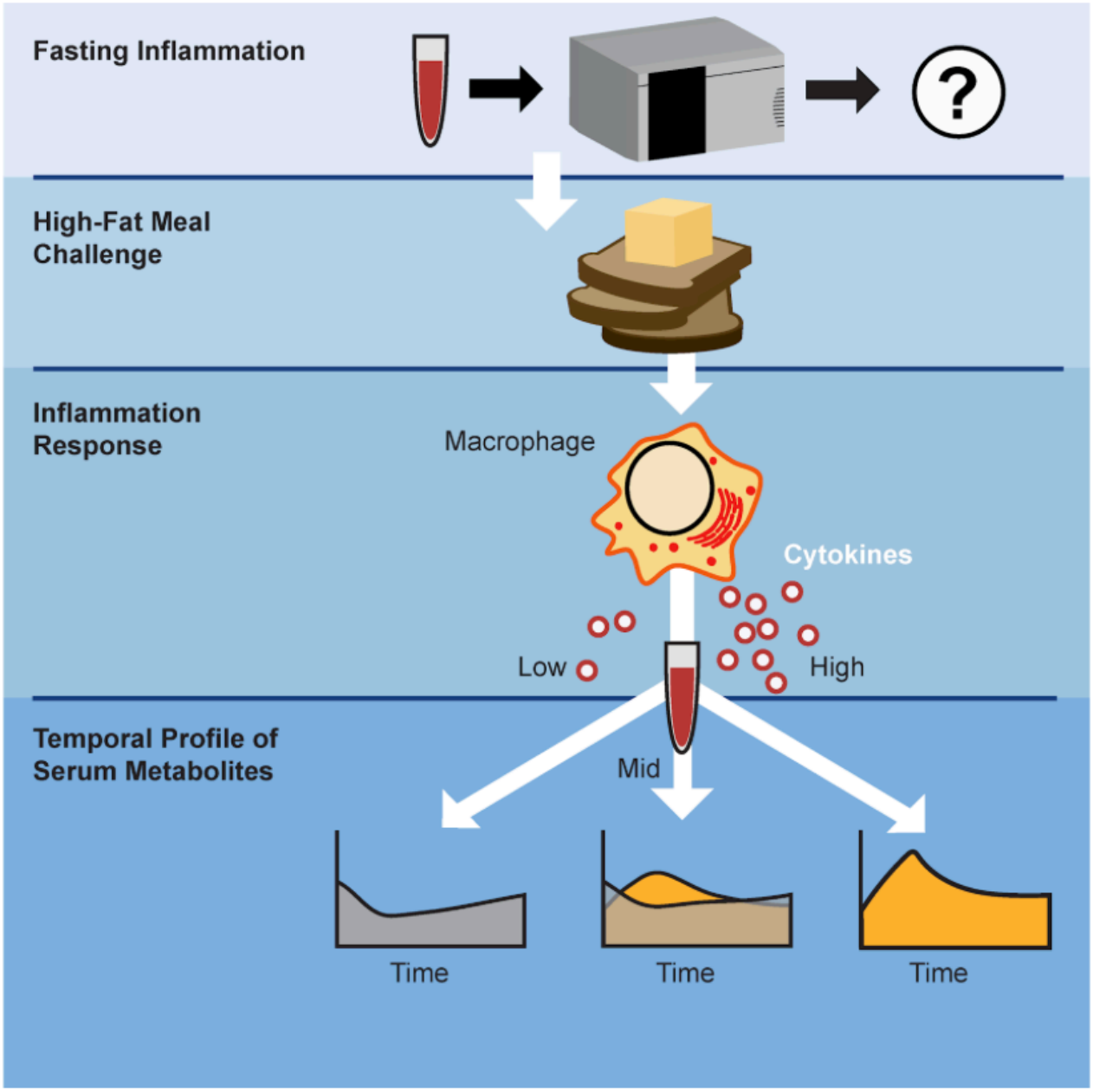


Figure 4.0. Graphical abstract.

Abstract

ABSTRACT-Chronic low-grade inflammation is a subclinical condition directly and indirectly linked to the development of a wide range of highly prevalent diseases responsible for the vast majority of morbidity. To examine mechanisms coupled to chronic disease, a group of overweight and obese human subjects without known inflammatory diseases were selected. Study participants underwent a high-fat meal challenge as an acute inflammation stimulus. Fasting and postprandial serum samples were analyzed for pro-inflammatory cytokines and metabolites. Analysis of serum metabolites grouped by baseline cytokine levels revealed that single samples had little power in differentiating the groups. However, an analysis that incorporated temporal response separated inflammatory response phenotypes and allowed us to create a metabolic signature of inflammation. Dynamic biosignatures were determined for all cytokines investigated and a composite model was created that incorporates three of the most commonly seen metabolites from the cytokine models: creatine, hydroxymethyluracil and histindinylhistidine. The models and the associated metabolites show the complexity of the inflammation response as well as metabolic components that are crucial to a cytokine-mediated inflammation response. The use of temporal response rather than absolute concentrations dramatically improved clinical diagnostics. This outcome led to the development of a dynamic biosignature as a powerful tool for stratifying risk to a wide range of diseases.

Introduction

Inflammatory diseases including type 2 diabetes, ischemic heart disease, neurodegenerative diseases, and cancer represent a group of human illnesses increasing in

prevalence and causing the bulk of human deaths worldwide (Tate et al., 2009, Katakami, 2018, Denver and McClean, 2018, Park et al., 2018). While there are many causes of low-grade inflammation, one of the most common is upregulation of inflammatory activity in response to hypertrophy-induced adipose tissue dysfunction, particularly visceral adipose tissue (VAT) located within the abdominal compartment (Ringseis, et al., 2015, Engin, 2017). Thus, low-grade inflammation is a putative driver of obesity related disease. However, a substantial proportion of obese individuals are metabolically healthy in that they do not meet the criteria for metabolic syndrome (Denis, et al., 2017, Bagheri, et al., 2018). The extent to which low-grade chronic inflammation alters metabolism in overweight and obese individuals is unclear. Elucidation of the metabolic impacts of being overweight or obese with and without inflammation is needed.

Metabolic impacts of inflammation stem from the actions of inflammatory cytokines and mediators on metabolic pathways (Gregor and Hotamisligil, 2011, Hotamisligil, 2017). Low-grade inflammation is characterized by persistent elevation in pro-inflammatory markers stemming from increased abundance of classically activated M1 macrophages, particularly the cytokines tumor necrosis factor (TNF)-α, interleukin (IL)-1β and IL-6 (Hotamisligil, 2017). These cytokines have impacts on nutrient metabolism, including the impairment of glucose metabolism. However, impacts of nutrients on inflammation have also been identified, creating a complex web of interactions that is difficult to assess from static biomarkers.

Metabolomic based biomarkers have been shown to be indicators of disease states while also providing insight into mechanistic causes of disease (Soga, et al., 2006, Badawi, et al., 2010, Considine, 2019). Biomarkers have been identified for many diseases, including ones with a known inflammatory component such as diabetes and atherosclerosis (Stenvinkel, et al., 1999,

Wheellock, et al., 2013). Patients with chronic low-grade inflammation have different metabolomic profiles that correlate with inflammation status (Pietzner, et al., 2017, Fitzpatrick and Young, 2013). Although biomarkers for chronic low-grade inflammation have been reported, biomarkers for progression from chronic low-grade inflammation to inflammatory diseases are limited. This may be due to the challenge of tracking temporal biological processes, with a static, single time-point measurement (Tokmina-Lukaszewska, et al., 2014). The purpose of this study is to compare static and temporal metabolomic approaches for predicting the magnitude of inflammatory response and then to elucidate metabolic impacts of those inflammatory responses (**Figure 4.0**). The power of our approach arises from looking at metabolite levels over time leading to the generation of a 'dynamic biosignature'.

Materials and Methods

Ethics statement

The protocol was approved by the Institutional Review Board at Montana State University. Written informed consent was obtained from all participants prior to their participation. The IRB protocol number is MM021116-FC.

Participants

Forty overweight and obese (BMI >25), although otherwise apparently healthy men and women participated in testing of anthropometric and metabolic markers and ingestion of a 50g high-fat meal challenge. Participants were 18-55 years old within a body mass index (BMI) of 27 - 36 kg/m^2. Potential participants were excluded if they had taken oral antibiotics within 90 days of study enrollment, regularly used anti-inflammatory medications, used estrogen-only

contraceptive, had a wheat and/or dairy allergies, were pregnant, or had any musculoskeletal, cardiovascular, gastrointestinal, or immunological condition that could interfere with the study. (clinicaltrials.gov #NCT04128839)

Anthropometrics

Measurements were collected from participants using the validated segmental multifrequency bioelectrical impedance analysis (SECA mBCA 515, Hamburg, Germany) (Bosy-Westphal). Fat mass (%) and estimated visceral adipose (L) were used for analysis.

High-fat meal challenge

The high-fat meal contained salted butter (58.3 g, Tillamook) over 3 pieces of whole wheat toast (127.5 g; Wheat Montana). Total energy content of the meal was 714 kcal, with 43.1% from fat, with a macronutrient breakdown of 50 g fat, 54 g carbohydrate, and 12 g protein. Water was provided with the meal; caffeinated black tea was provided for participants who identified as habitual coffee consumers.

Blood sampling

Participants were instructed to avoid alcohol consumption and strenuous physical activity in the 24 hours prior to their visit and to complete an overnight fast (10 - 12 hours) before blood collection. Participant blood samples were collected by a certified nurse or physician in the morning before ingestion of the meal and hourly for 4 hours after meal ingestion, totaling five time points. Whole blood in serum separating tubes was allowed to clot for 15 minutes before centrifugation at 1200 RPM for 15 minutes with resulting serum aliquoted and stored at -80ºC until analysis.

Determination of blood markers

Blood markers of metabolic syndrome were determined from whole blood run on Picollo Xpress Chemistry Analyzer lipid panels (Abaxis, Union City, USA). Cytokine measurement was performed using high-sensitivity multiplexing technology (Bio-Rad Bio-Plex® 200 HTS) following procedures by Millipore (EMD Millipore Corporation, Billerica, USA). Selected pro-inflammatory cytokines include granulocyte macrophage colony stimulating factor (GM-CSF), four interleukins (IL) including IL-1β, IL-6, IL-17, IL-23, and tumor necrosis factor alpha (TNF-α). Serum samples at each time point during the high-fat meal challenge were run in duplicate.

Low-grade inflammation and inflammation response group assignment

Assignment of participants to low, mid or high low-grade inflammation groups was made based on a k-means analysis of the cytokine responses (Hartigan 1979). K-means analysis is an unsupervised machine learning algorithm that groups data in the most significant way possible while trying to maintain the lowest number of groups. To start the analysis, the cytokine response first needed to be determined. Cytokine response was calculated by taking the time point with the highest post-prandial concentration for each cytokine and then subtracted this value from the basal cytokine concentration. Response values were determined for each of the six cytokines using this method. Cytokine responses were then examined using a k-means analysis in R. The results of the k-means analysis clustered the 40 subjects into three inflammation groups that correspond to low, medium and high inflammation response. 17 subjects were included in the low inflammation response group, 14 were included in the medium inflammation response group and 9 were included in the high inflammation response group.

Composite response scores were also created by normalizing values between cytokines to give each cytokine an equal contribution to the overall response score.

Metabolite extraction

Frozen serum samples were thawed and 20µL was placed in a clean tube. 80µL of HPLC grade methanol was added to the sample after which it was vortexed briefly and placed in a -80°C freezer for two hours. After two hours, the sample was centrifuged at 20,000g for 10 minutes. The metabolite supernatant was collected and concentrated in a Speed Vac to dryness while the protein pellet was discarded. Samples were then stored at -80°C until ready for liquid chromatography mass spectrometry (LCMS) analysis at which time they were reconstituted with 40µL of methanol:water (50:50) and placed in a clean mass spectrometry vial.

LCMS conditions

Analysis was completed on an Agilent 6538 Q-TOF MS coupled to an Agilent 1290 UHPLC using a 130Å, 1.7µm, 2.1mm X 100mm Acquity BEH-HILIC HPLC column. Samples were ionized via electrospray ionization and runs were completed in positive mode. Mobile phase A was 15mmol/L ammonium formate and mobile phase B was ACN using a 10-40% A gradient over 6 minutes. Flow was kept at 400µL/minute and the column compartment temperature was set at 30ºC (Awwad, et al., 2016). Pooled serum samples were included in runs before, during and after sample analysis to ensure LCMS function and repeatability of analysis. MSMS analysis, or tandem mass spectrometry, was completed using the same LC conditions while targeting specific ions using retention time and m/z values from previous MS runs.

Data analysis

After LCMS analysis completion, raw data files were converted to .mzML files using MSConvert (Chambers, et al., 2012). Vendor peak picking was used along with a threshold of the 300 most intense ions per scan for data conversion. Data was then mined with mzMine using an intensity minimum value of 1,000 counts based on a visual inspection of the total ion chromatogram to remove noise (Katajamaa, et al., 2006). An error of 20 ppm was used with mzMine along with a retention time window of 0.2 minutes to differentiate unique peaks with similar m/z values. Retention times were analyzed for creatine, m/z value 114.065, to determine retention time drift over the course of the LCMS analysis (Supplemental Table 2). Blank samples were also run and the resulting features were removed from the biological data if present at a ratio under 5:1 in the sample compared to the blank. Mined data were then input to MetaboAnalyst for statistical analysis including an ANOVA analysis. Sample data was not filtered in MetaboAnalyst due to the number of features and data was normalized using the auto scaling function. Multiple testing correction was accomplished using Tukey's test. Reproducibility was confirmed by determining the coefficient of variation for each selected metabolite's ion intensity from the pooled samples that were ran before, during and after the experimental analysis (Supplemental Table 3). MSMS data were converted to .mgf files using MSConvert and analyzed with Sirius software to obtain compound identifications (Dührkop, et al., 2015, Dührkop, et al., 2019). Molecular formula and molecule identifications were determined by searching the Human Metabolomics Database (HMDB). SIRIUS was used to generate formulas while CSI:Finger ID was used for molecule identification. Both searches were done within a 20ppm error window. Reported identifications were the top metabolite selected by

CSI:FingerID. The ZODIAC network tool was used to confirm chemical formula selection and CANOPUS software was utilized to determine the compound class of unknown features.

Data and software availability

The mass spectrometry, participant and validation tables have been deposited in Mendeley Data: Miles, Mary (2021), "Temporal metabolic response yields a dynamic biosignature of inflammation. Peach et al.," Mendeley Data, V1, DOI URL: https://data.mendeley.com/datasets/j9rfpwbjhj/1.

Results

Participants with a BMI >27 and < 35 $kg \cdot m^{-2}$ and no overt health concerns were screened for fasting glucose, inflammation cytokine concentration and other factors before consumption of a high-fat meal, which is an established test to induce inflammation in humans (Herieka 2014). Subsequent glucose and cytokine concentrations were determined at 1, 2, 3, and 4 hours post-consumption. Inflammation was determined using a high sensitivity cytokine panel that included TNFα, GM-CSF, IL-1B, IL-6, IL-17 and IL-23**.** The maximum difference between basal and post meal value for each cytokine over the five time points was used to generate an inflammation response value for each participant. A comprehensive response value was then calculated for each participant by summing the normalized response values for each cytokine. The comprehensive response values were then grouped using a k-means analysis and participants were designated as either high- mid- or low-inflammation responders based on the magnitude of their inflammation response (Hartigan 1979) **(Figure 4.1A).** Seventeen subjects were classified as low-responders, fourteen as mid-responders and nine as high-responders **(Figure 4.1B,**

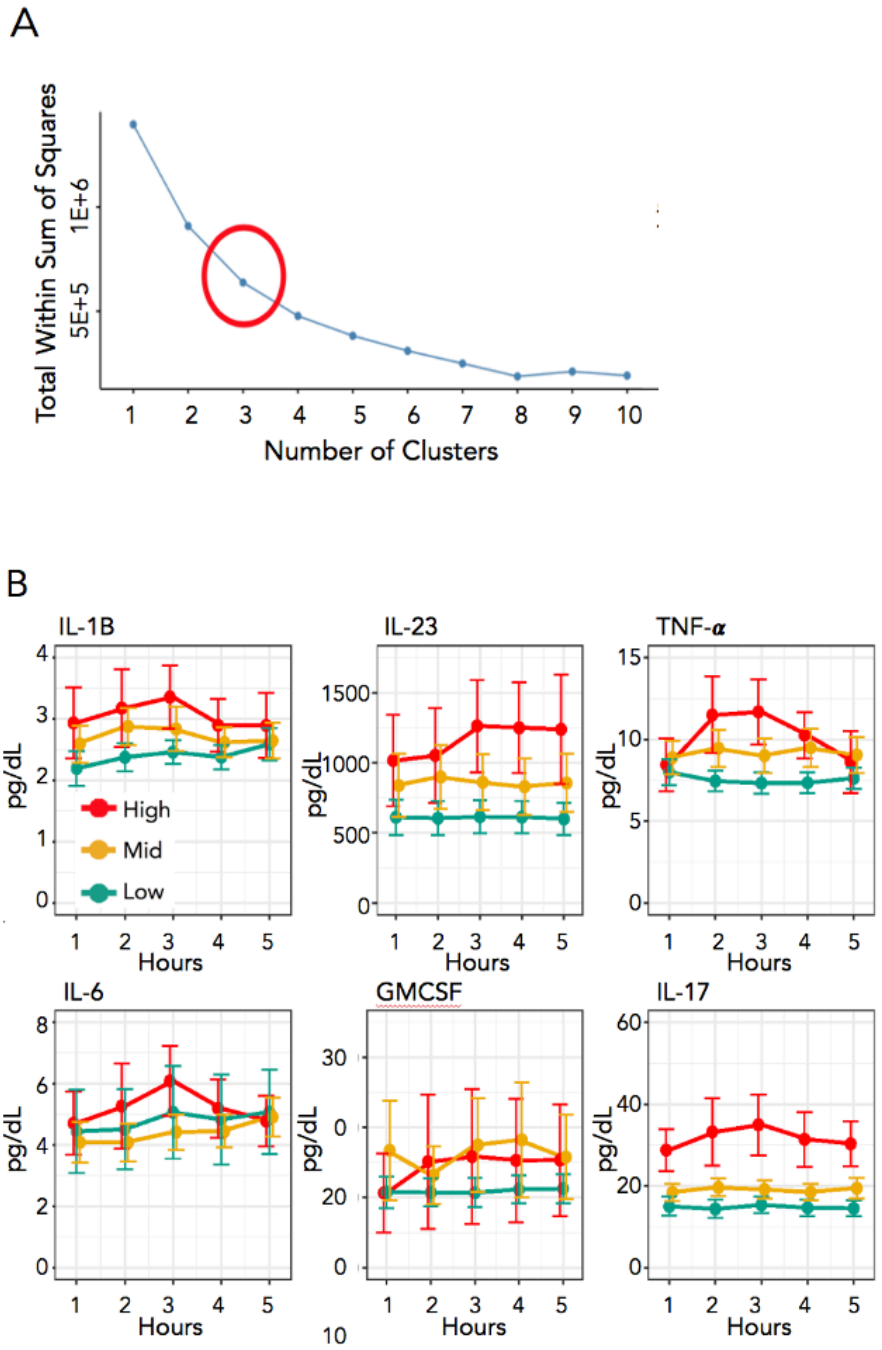


Figure 4.1. Grouping of participants by cytokine response. (A) K-means analysis of the cytokine concentrations for participants in the cohort. The "elbow" of the optimal clusters graph shows that three groups is the best way to bin the subjects based on cytokine response. (B) Plots for pro-inflammatory cytokines used in the study at each time point from fasting to four hours post meal.

Supplemental Table 4.1). High- and mid-responders had lower fasting cholesterol than low-responders but did not vary in other measured anthropomorphic or static metabolic characteristics. To generate a metabolic profile for each participant, a non-targeted metabolomic analysis of serum samples was conducted using LCMS. An established HILIC method facilitated rapid, reproducible, and deep sample profiling (Awwad, et al., 2016). Raw data was retention-time aligned and background signals were removed using mzMine (Katajamaa, et al., 2006, Patti, et al., 2019). Through this process a dataset of 560 unique mass features was obtained. The high-resolution LCMS data showed that the overall metabolic state was similar across the participant pool and did not change dramatically for a given individual during the time series (**Supplemental Figure 4.1**).

The composite cytokine response k-means grouping was used to analyze the LCMS at each time-point using MetaboAnalyst (Chong, et al., 2019). Principal component analysis (PCA) was performed to examine the separation between inflammation response groups and the impact of time. A PCA analysis of response groupings showed extensive overlap of the groups (**Figure 4.2A, Supplemental Figure 4.2**). ANOVA analysis confirmed the groups could not be differentiated, as no metabolites were significantly different at a p-value of <0.05 across the time series. Metabolomic profiles were examined in closer detail by generating heatmaps, which showed little to no clustering of specific metabolites based on cytokine-based inflammation responses (**Figure 4.2B**). Our conclusion is that these statistical analyses have limited power to differentiate inflammatory responders at specific time points based on the intensity of metabolic features.

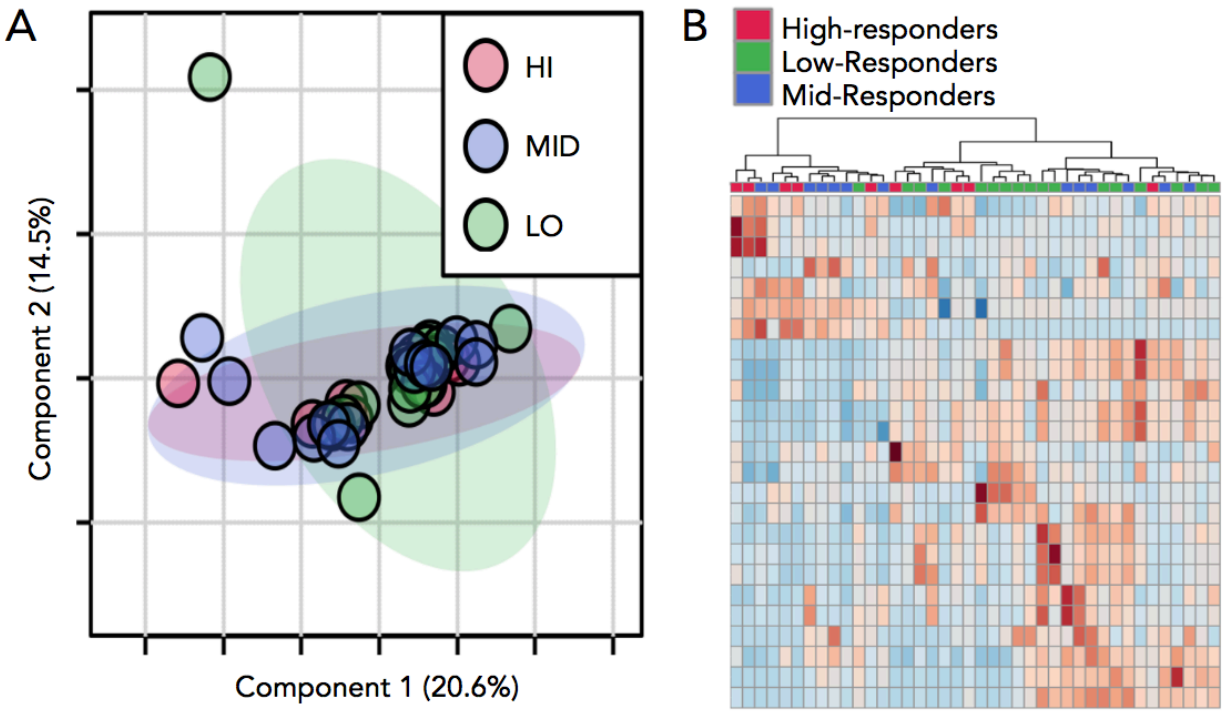


Figure 4.2. Static metabolic profiling of high, mid and low-inflammation groups. (A) A principal component analysis (PCA) of response groups at one-hour post-meal show little to no separation. (B) Heatmap for time point two showing the top 25 discriminating features. The dendrogram at the top of the figure shows little clustering of the inflammation groups.

Inflammatory response is a time-dependent process; therefore, we reasoned that even quantitative metabolomics data based on a single time point may fail to capture objective differences (Grigorov, 2001). This is particularly challenging with human subjects as there is a wide range of biological and lifestyle variation. To address this, we analyzed the metabolomic profiles as dynamic datasets with the addition of a temporal component to match our cytokine response groupings. Using the time-series software in MetaboAnalyst, an analysis of variance-simultaneous component analysis (ASCA) was performed to identify features which contribute to systemic variation in a time-resolved manner. ASCA analysis is accomplished by producing interaction matrices for each feature and then determining the contribution of each feature to systemic variation, much like a PCA model (Saccenti, et al., 2018, Smilde, et al., 2005). This method allows for analysis of multivariate time-course data. Discriminating temporal features were then filtered using a squared prediction error (SPE) and a leverage calculation (**Figure 4.3A**). This yielded 17 well-modeled features based on q-values and leverage criteria that have group specific concentration patterns.

The next step was to identify the 17 selected features. This was completed using MSMS to generate collision induced fragmentation patterns which were matched using an in-house standards library and SIRIUS formula and structure prediction software (Dührkop, et al., 2015, Dührkop, et al., 2019). Based on high resolution mass measurement, retention time, and fragmentation pattern, chemical formula could be assigned for all 17 features and confident identifications were made for 13 metabolites (**Table 1**). Identification of the ions at m/z 133.131, 164.142, 240.151 and 82.026 were inconclusive based on available fragmentation databases.

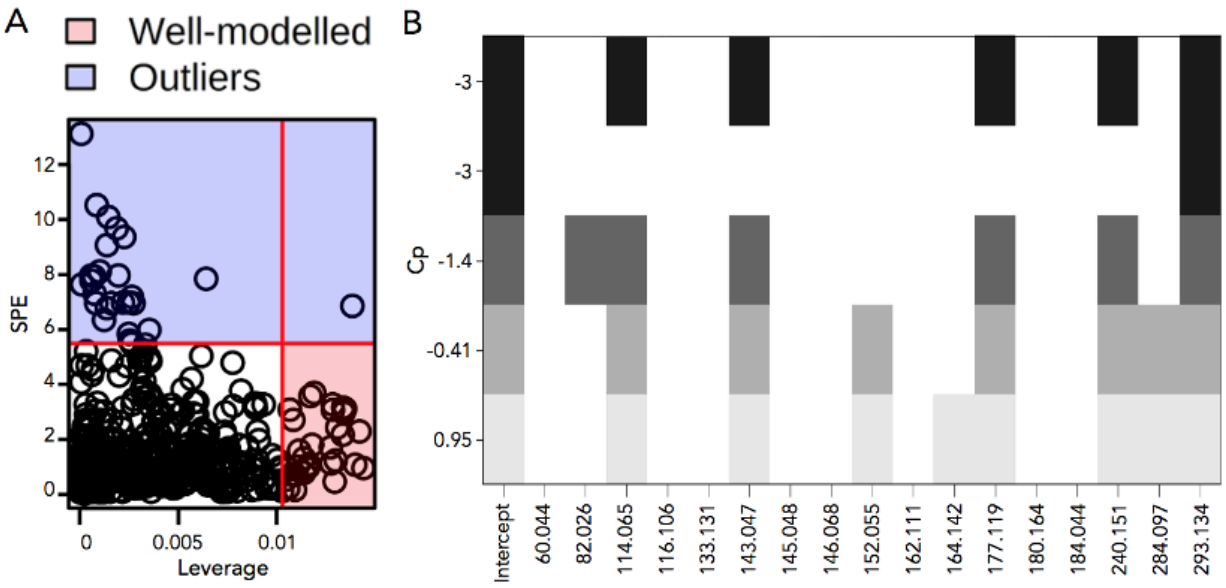


Figure 4.3. Feature interaction based on time and inflammation response. (A) Important features for the interaction between response groups over the time-course found by combining SPE scores and leverage. (B) Cp plot of well-modeled features in a linear regression model predicting cytokine response. Columns represent mass features and rows show BIC values where the best model is on the top row.

Table 4.1. Identification of important metabolites

Metabolite	Mass	RT	Validation
Acetamide	60.044	1.4	Authentic Standard
C4H3NO	82.026	2.4	MSMS
Creatine	114.065	1.8	Authentic Standard
Proline	116.106	3.5	Authentic Standard
Organonitrogen compound C6H16N2O	133.131	5.0	MSMS
Hydroxymethyluracil	143.047	1.8	MSMS
Threitol	145.048	1.5	Authentic Standard
4-Guanidinobutyrate	146.068	2.4	MSMS
Guanine	152.055	1.5	Authentic Standard
Carnitine	162.111	3.5	Authentic Standard
Aralkylamin compound C11H17N	164.142	3.4	MSMS
(5-Amino-6-hydroxyhexyl)carbamate	177.119	4.9	MSMS
Tigylglycine	180.064	1.7	MSMS
Butyrylglycine	184.044	1.2	MSMS
Azacyclic compound C10H17N5O2	240.151	1.3	MSMS
Guanosine	284.097	1.5	Authentic Standard
Histidinylhistidine	293.134	3.5	MSMS

Linear models were then generated and tested using the 17 selected metabolites to describe the pro-inflammatory cytokine baseline to peak response values and models with the best Mallow's Cp values (**Figure 4.3B**) were determined (**Table 4.2)**. Linear models were created for each of the six individual cytokine responses that included eleven of the selected metabolites, nine of which were successfully identified (**Table 4.3**). Four of the six models, corresponding to IL1B, IL6, TNFα and GMCSF response, were significant to a p-value of <0.05, while the IL17 and IL23 response models had p-values of 0.12 and 0.068, respectively. The models associated with IL1B, IL6, TNFα and GMCSF had the highest coverage as well, with R^2 values between 0.65 to 0.25. In contrast, the models for IL17 and IL23 had R^2 values below 0.1. Along with individual cytokine models, a composite model was created to better describe the inflammation response as a whole and was composed of five metabolites. A key point is that the cytokine models had the power to predict inflammation response phenotype using only serum metabolite data.

To validate the dynamic biosignature, data from two independent trials were obtained. Participants were part of food supplementation studies that included a high-fat meal challenge, metabolomic profile analysis, and pro-inflammatory cytokine determination. The only criteria for selection was having a BMI >27 and < 35 $kg \cdot m^{-2}$. Each subject was assigned to the appropriate cytokine inflammation group and metabolites in the composite dynamic biosignature were compiled and placed in the composite linear model. For the 13 participants, the metabolic biosignature assigned 9 to the correct inflammation group (**Table 4.4)**. Three of the four mis-assignments involved the mid-inflammation designation. Interestingly, the remaining participant

Table 4.2. Matrix showing the metabolites in each model

	Model						
Metabolite	IL1B	IL6	IL17	IL23	TNF	GMCSF	Composite
Creatine	X				X	X	X
Hydroxymethyluracil	X	X		X			X
Histidinylhistidine	X				X	X	X
Threitol	X					X	
4-Guanidinobutyric acid	X					X	
Butyrylglycine			X			X	
(5-amino-6-hydroxyhexyl)carbamic acid					X		X
240.151					X		X
Guanine		X					
Guanosine		X					
82.026					X		

Table 4.3. Models for each cytokine response

Cytokine	Metabolites	p.value	R2
IL1B	Creatine + Hydroxymethyluracil + Threitol + 4-Guanidinobutyrate + Histidinylhistidine	1.40e-02	0.330
IL6	Hydroxymethyluracil + Guanine + Guanosine	1.60e-02	0.250
IL17	Butyrylglycine	1.20e-01	0.062
IL23	Hydroxymethyluracil	6.80e-02	0.085
TNF	m/z 82.026 + Creatine + (5-amino6-hydroxyhexyl)carbamic acid + m/z 240.151 + Histidinylhistidine	9.50e-05	0.520
GMCSF	Creatine + Threitol + Butyrylglycine + 4-Guanidobutyrate + Histidinylhistidine	6.21e-07	0.650
Composite	Creatine + Hydroxymethyluracil + (5-amino6-hydroxyhexyl)carbamic acid + m/z 240.151 + Histidinylhistidine	1.90e-02	0.320

Table 4.4. Validation study

Participant	Study	Predicted	Acutal
1*	Lentil	LO	MID
2	Lentil	LO	LO
3	Lentil	LO	LO
4	Lentil	HI	HI
5	Lentil	LO	LO
6	Lentil	LO	LO
7	Lentil	HI	HI
8*	Lentil	MID	LO
9*	Lentil	LO	HI
1	Aronia	LO	LO
2	Aronia	LO	LO
3*	Aronia	HI	MID
4	Aronia	LO	LO

An * indicates participants who were incorrectly predicted

had a dynamic biosignature that placed them in the low inflammation group, while their cytokine profile was consistent with high.

A closer investigation of the model shows that between the basal and one-hour post high-fat meal challenge, the high- and low-inflammation response groups have an inverse slope (**Figures 4.4-5**). This diametric relationship between high- and low-inflammation groups was consistent in 8 of the 11 features included in the individual and composite models. Of the eight features with this pattern, the mid-inflammation response group trends with the high-response group for seven of the metabolites and the low-response group for the unknown metabolite m/z 240.151. We describe this and similar patterns of differing metabolic responses between inflammation groups over time as a dynamic biosignature. To understand the metabolic impacts of concentration changes in the selected metabolites, we further examined each metabolite in the context of their metabolic roles and specific patterns in the time-series.

Discussion

A major finding of this study was that temporal metabolomic responses to an inflammatory stimulus generated clear physiological distinctions not present under basal conditions that were predictive of cytokine-based inflammation responses. A traditional static biosignature can be an effective indicator of a specific metabolic activity or disease state. However, in our case a static biosignature was unable to convincingly differentiate subjects into groups based on inflammation phenotype. Our breakthrough is the demonstration that the change in metabolic state over time is a more powerful way to differentiate underlying physiologic states than measuring absolute values at any given time. This is analogous to the role of derivatives in

the analysis of mathematical functions. Values are based on the change in concentration over a change in time and not on absolute initial or final concentration.

Data from 560 serum metabolites collected before and at four time points after a high-fat meal challenge using LCMS, were used to construct a predictive model. By coupling the change in markers of inflammation and response kinetics, metabolites having a significant difference between inflammation groups were discovered. Using selected metabolites, linear models were created for each cytokine and a composite cytokine model was also created. Models showed overlap between metabolites with several, including histidinylhistidine and creatine, belonging to the same cytokine model (**Table 4.3**). This suggests involvement in the same metabolic response, i.e. histidinylhistidine and creatine abundance are linked. This is consistent with known metabolic pathways and validates the approach used to generate the models (Smith and Ning, 1961).

A further investigation of modeled metabolites from inflammation response groups showed distinct and repeated patterns between the basal and two-hours post-meal metabolomic profiles which includes three time-points (**Figure 4.4-5**). For each metabolite in the composite model, an initial opposing slope direction between the basal and one-hour post-meal concentration is observed, after which metabolites show a general trend back to the basal concentration two hours post-meal. Three of the five metabolites demonstrated an increase in the first hour for the high-responders and a decrease in the first hour for the remaining two metabolites. Low-responders displayed the opposite trend. The difference in the concentration changes of the modeled metabolites show that an important temporal component of the cytokine immune response is correlated to the change in concentration of specific metabolites at one-hour

post-meal. However, this trend was not universal with the rest of the metabolites in the cytokine models. Of the six selected metabolites not in the composite model, four had the same pattern of inverse slopes between high- and low-responders from basal to one-hour post-meal seen in the composite model metabolites. Modeled metabolites were investigated for known associations with cytokine specific inflammation, inflammation in general, or an inflammatory disease. Several of the metabolites fit into one or all of these categories with some having links between the metabolite and the cytokine in a specific model.

Composite model metabolites

The composite model included creatine, hydroxymethyluracil, histidinylhistdine, (5-amino-6-hydroxyhexyl) carbamic acid and an unknown metabolite with an m/z value of 240.151.

Creatine is part of the composite model as well as three individual cytokine models. Creatine is synthesized from amino acids and can be converted to phosphocreatine which is used to replenish ATP stores in muscle cells through the action of creatine kinase (Wallimann, et al., 2011). In our cohort, creatine was included in the IL1B, TNFα, GMCSF and composite models. Previous work has shown that IL1B and TNFα are inversely correlated with creatine concentration (Bassit, et al., 2010). IL1B and TNFα are both alarm cytokines that are expressed early in an inflammatory response due to increases in stress hormones and promote production of other pro-inflammatory cytokines (Baumann and Glaudie 1994, Cupps and Fauci 1982). Phosphocreatine and creatine help to maintain ATP availability and barrier function in the intestinal mucosa during inflammation. High-fat and sugar intake induces increased creatine uptake in the intestinal epithelium (Colgan, et al., 2015, Poupin, et al., 2019). Low-responders showed an initial increase in creatine concentrations followed by a return to fasting

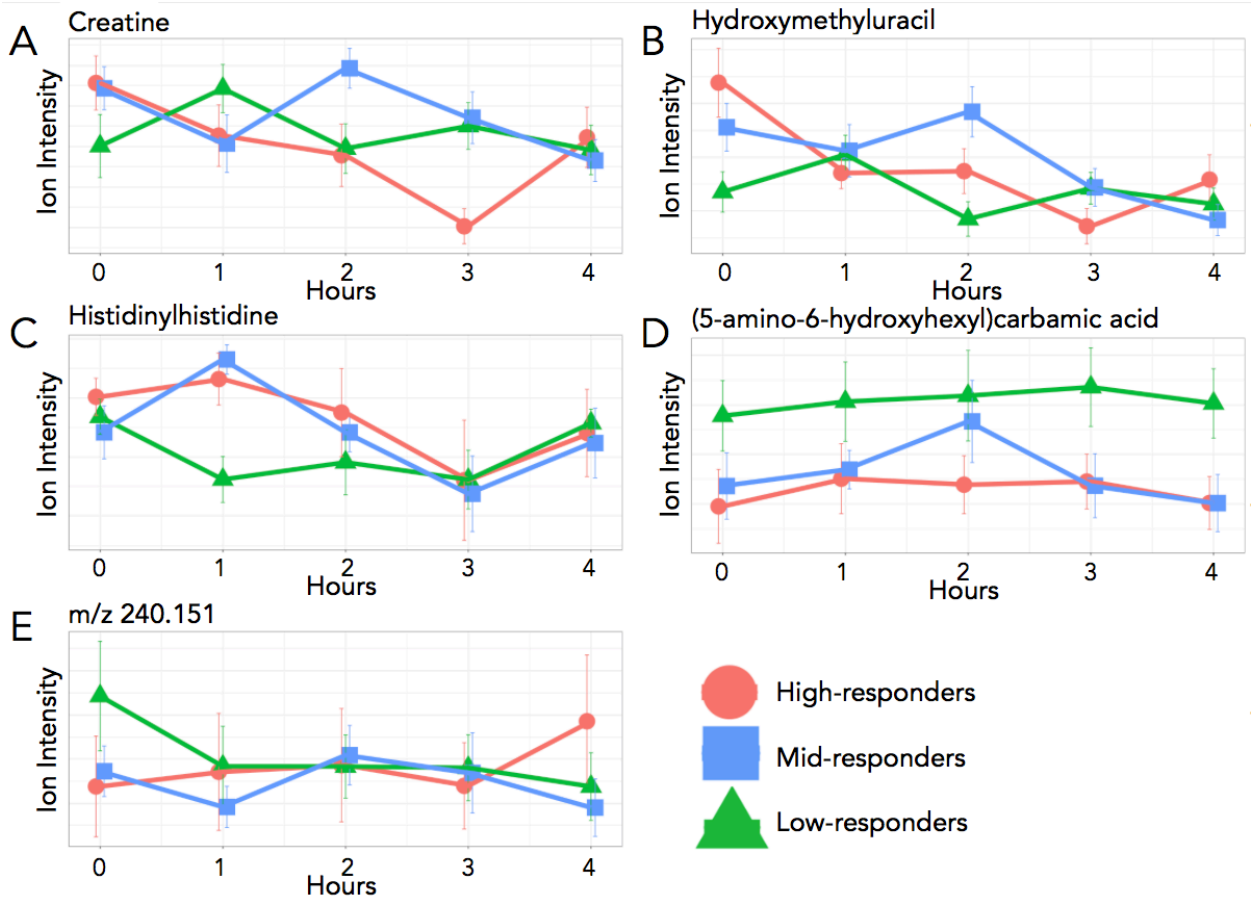


Figure 4.4. Composite model. Plots of the five metabolites found in the composite model including error bars. Relative concentrations are shown for the entirety of the time course. Creatine, hydroxymethyluracil and histidinylhistidine represent the three most common metabolites, found in three cytokine models and the composite model while carbamic acid and the unknown metabolite were found in one model each and the composite model. Note the slope change between the high- and low-response groups at one-hour post-meal.

concentrations after two hours (**Figure 4.4A**). Mid- and high-responders had an opposite trend during the first hour showing increasing creatine concentrations with an eventual return to below fasting levels. Our data is consistent with IL1B and TNFα being early responders to inflammation as well as being inversely related to creatine concentration.

Along with creatine, hydroxymethyluracil was included in the composite model as well as three individual cytokine models. Hydroxymethyluracil is a uracil derivative and is a biomarker for oxidative stress and DNA damage (Decarroz, et al., 1986). It has also been shown to be more abundant in disease states such as in colorectal cancer (Rozalski, et al., 2015). Hydroxymethyluracil fit into models for IL1B, IL6, IL23 and the cytokine composite model. In a recent study, the removal of thymine DNA glycosylase, a DNA oxidation repair mechanism, resulted in the increase of hydroxymethyluracil and IL1B (Tricarico, et al., 2019). Alternatively, increased intracellular ascorbic acid (vitamin C) has been shown to induce hydroxylation of thymine leading to an increase of intracellular hydroxymethyluracil in cell culture (Modrzejewska, et al., 2016). Our data shows that the high- and mid-responders decreased in hydroxymethyluracil with the introduction of the high-fat meal while the low-responders saw an increase (**Figure 4.4B**). Thus, the elevation in hydroxymethyluracil may reflect adaptation of the gut to different nutritional environment between groups that could have conferred some level of resistance to inflammatory stimuli in the low-responders. This is a novel observation worthy of follow up with respect to the potential of antioxidant supplements to mitigate inflammation in the gut.

The final metabolite to be included in not only the composite model but three additional cytokine models is histidinylhistidine. Histidinylhistidine, also known as histidylhistidine, is

composed of two histidine residues and is a product of incomplete protein degradation. This dipeptide appeared in four cytokine models; IL1B, TNFα, GMCSF and the composite model. Both IL1B and TNFα are potent stimulators of proteolysis and GMCSF is a pro-inflammatory cytokine that stimulate proliferation and activation of phagocytic leukocytes capable of producing IL1B and TNFα (Chang and Bistrian, 1998, Shiomi and Usui, 2015). These are the same three cytokines models that creatine appears in, and there has been some research showing an effect of histidine on creatine concentrations leading us to postulate that histidinylhistidine may be involved in creatine metabolism signaling (Smith 1961). If so, this effect would most likely be inhibitory. A clear inverse relationship is seen within response groups between creatine and histidinylhistidine. Although an association has been observed, this avenue of investigation does not appear to have not been explored since 1961.

One-hour post-meal, high- and mid-responders have an increase in histidinylhistidine corresponding to a decrease in creatine during the same time period (**Figure 4.4C**). Low-responders have an initial decrease in histidinylhistidine and an increase in creatine in the first hour. Histidinylhistidine also has a link to human carbonic anhydrase III (HCA III), a metalloenzyme that converts CO_2 to HCO_3. This reaction is completed in a two-part process consisting of the conversion of CO_2 to HCO_3 followed by the release of HCO_3 and its replacement with water. This reaction is significantly faster in other human carbonic anhydrases, such as HCA II, and appears to be a result of a residue substitution at Phe 198. This does not hinder the first portion of the reaction, but sterically inhibits binding of proton rich molecules needed to facilitate the release of HCO_3 from the enzyme. Histidinylhistidine has been found to bind to Phe 198 and activate the release of HCO_3 (Duda, et al., 2005). Changes in HCO_3

concentrations have an effect on inflammatory responses through the modulation of cytokine expression including a direct relationship with several interleukins (Ori, et al., 2015, Kawakami, et al., 2020). Therefore, histidinylhistidine binding to Phe 198 on HCA III leads to an increase in HCO_3 concentrations, which in turn causes an increase in interleukins and an increased inflammation response.

The fourth metabolite in the composite model is (5-amino-6-hydroxyhexyl) carbamic acid. Carbamic acids are involved in numerous metabolic processes including the transport of carbon dioxide in hemoglobin. Although there are currently no links to inflammation, this metabolite should be investigated further. Concentrations differed between the low-inflammation and high and mid-inflammation responders from two to three hours with low-responders increasing and mid- and high-responders decreasing (**Figure 4.4D**). Carbamic acid was also part of the TNFα model.

The final metabolite to be included in the composite model was unable to be identified using our standard library or MSMS fragmentation pattern matching. This unknown has a m/z value of 240.151 ±0.005. Although not identified, the unknown metabolite could be an integral part of the dynamic biosignature for inflammation as one-hour post-meal shows two different trends (**Figure 4.4E**). Low- and mid-responders decreased in concentration while high-responders increased. This metabolite was also included in the TNFα model.

<u>Cytokine specific model metabolites</u>

Of the eleven metabolites selected for models, six of them were not included in the composite model. Three metabolites; glutamine, threitol and 4-guanidinobutyric acid, were part of two different models. The final three metabolites; guanine, guanosine and the unknown

feature with an m/z value of 82.026, were each included in just one model. However, even though these metabolites were not included in the composite model or in a large number of cytokine models, it does not mean that they do not have an important impact on the inflammatory response.

4-guanidinobutyrate is a product of arginine metabolism and the first of the metabolites included in two cytokine models. Inflammatory stimuli indirectly effect 4-guanidinobutyric acid production by upregulation nitric oxide production, which in turn inhibits production of 4-guanidinobutyric acid (Satriano 2004). Increases in 4-guanidinobutyrate were also seen in human monocytes infected with Lyme disease (Kerstholt, et al., 2018). Corresponding to the increase in 4-guanidinobutyric acid in the Lyme disease infected monocytes was an increase in pro-inflammatory cytokines. 4-guanidinobutyrate was part of the IL1B and GMCSF models. We measured higher fasting 4-guanidinobutyrate in the low-responder group which decreased postprandially (**Figure 4.5A**). This shows that 4-guanidinobutyric acid may be part of an initial inflammation response or part of an early effect from the initial inflammation response

Threitol was also included in two cytokine models. Threitol is a product of xylose metabolism and has been associated with several inflammatory diseases including diabetes (Jing and Chengji. 2019). Threitol was part of our IL1B and GMCSF models and has been linked to IL1B. IL1B and threitol have both shown to lead to arterial calcification by modulating vascular maintenance cells (Collett and Canfield, 2005, Sage, et al., 2010). In our model, threitol concentrations increased in the pivotal first hour in the high- and mid-responders while decreasing in the low-responders (**Figure 4.5B**). This pattern reinforces the parallel findings related to arterial calcification.

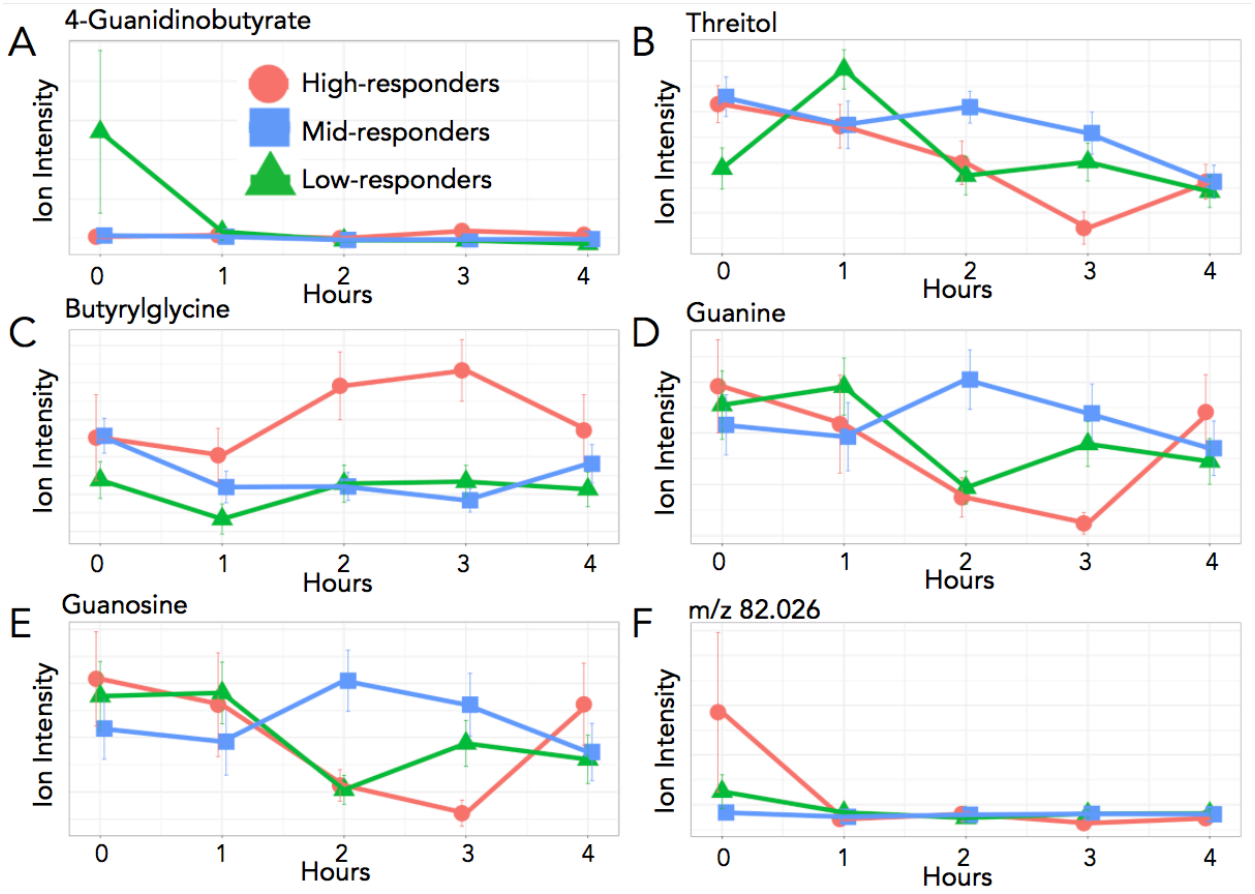


Figure 4.5. Six of the metabolites from the cytokine models. Each of these metabolites were found in specific cytokine models but were not included in the composite model. 4-Guanidinobutyrate, Butyrylglycine and Threitol were part of two cytokine models. The other metabolites were found in one model each. Relative concentrations are shown for the entirety of the time course with error bars.

The last metabolite selected for two models is butyrylglycine, an acyl glycine and a byproduct of butyryl-CoA accumulation. As fatty oxidation activity decreases, the initial compounds, like butyryl-CoA, increase in concentration and are degraded to compounds like butyrylglycine via various metabolic processes. Accumulation of short chain fatty acids is seen in individuals with short chain acyl-CoA dehydrogenase deficiency as short chain acyl-dehydrogenase (SCAD) catalyzes the first step in beta-oxidation for small fatty acids with four to six carbon chains (Lisyova, et al., 2018). Fatty acid oxidation rates have an inflammation specific outcome in macrophage mitochondria where increases in fatty acid oxidation, indirectly leading to decreases in butyrylglycine, have been indicated in a decreased inflammatory response (Namgaladze and Brüne, 2016). Butyrylglycine has also been shown to be differentially regulated in obese mice in response to stress associated with a differential inflammatory response (Haley, et al., 2017). Our cohort did not differentiate with respect to slope from the basal time to two hours post-meal. However, the high-response group continued to increase from two hours post-meal to three hours post-meal, contrasting the static trend seen in the low- and mid-responders over the same time period (**Figure 4.5C**). As differential processing of fatty acids from the high-fat meal challenge occurs over time, so would differential accumulation of butyryl-CoA, and therefore butyrylglycine. This could be due to many factors including SCAD differences in either effectiveness or concentration. Increased butyrylglycine would indicate a decrease in fatty acid oxidation and, if occurring in macrophages, a decrease in inflammation which aligns with our data. High-responders show a relatively large increase in butyrylglycine two hours post-meal while low-responders stay relatively stable. Fittingly, this metabolite was part of the GMCSF model, as well as the IL17 model.

The next two metabolites to be selected for a model were guanine and guanosine. Guanine is a nucleobase and a purine derivative while guanosine is a nucleoside comprised of a guanine attached to ribose. Guanine and guanosine have been shown to modulate T-cell mediated inflammation responses. They accomplish this via an anti-inflammatory effect by inhibiting cytokine production, including IL6 and IL2 (Shinohara and Tsukimoto, 2018). Guanine and guanosine were part of the IL6 model and both exhibit a pattern where low-responders increased in the first hour while high- and mid-responders decreased (**Figure 4.5D, 4.5E**). As an inhibitor of cytokine production, the decrease in the high- and mid-responders would be expected.

Finally, the unknown feature with an m/z value of 82.026 was also selected for one model. This feature was unable to be identified by either the in-house standard library or by MSMS. The unknow feature is in the TNFα model and our data shows a much higher fasting concentration in the high-responders group (**Figure 4.5F**). However, this concentration dramatically decreases relative to the other groups to such a degree that after one hour, the high-responders have a lower concentration than the other groups.

Taken together, these results paint a complex picture of dynamic inflammation response. Inclusion of metabolites in our models with ample support for roles in inflammation responses, such as creatine and hydroxymethyluracil, reinforce the importance of these metabolites in pro-inflammatory responses. By having well-documented connections between cytokines and specific metabolites in several models, credibility is lent to our statistical methodology and to our models. At the same time, this also provides support for the importance of the metabolites without corresponding inflammation associations in the literature, such as (5-amino-6-

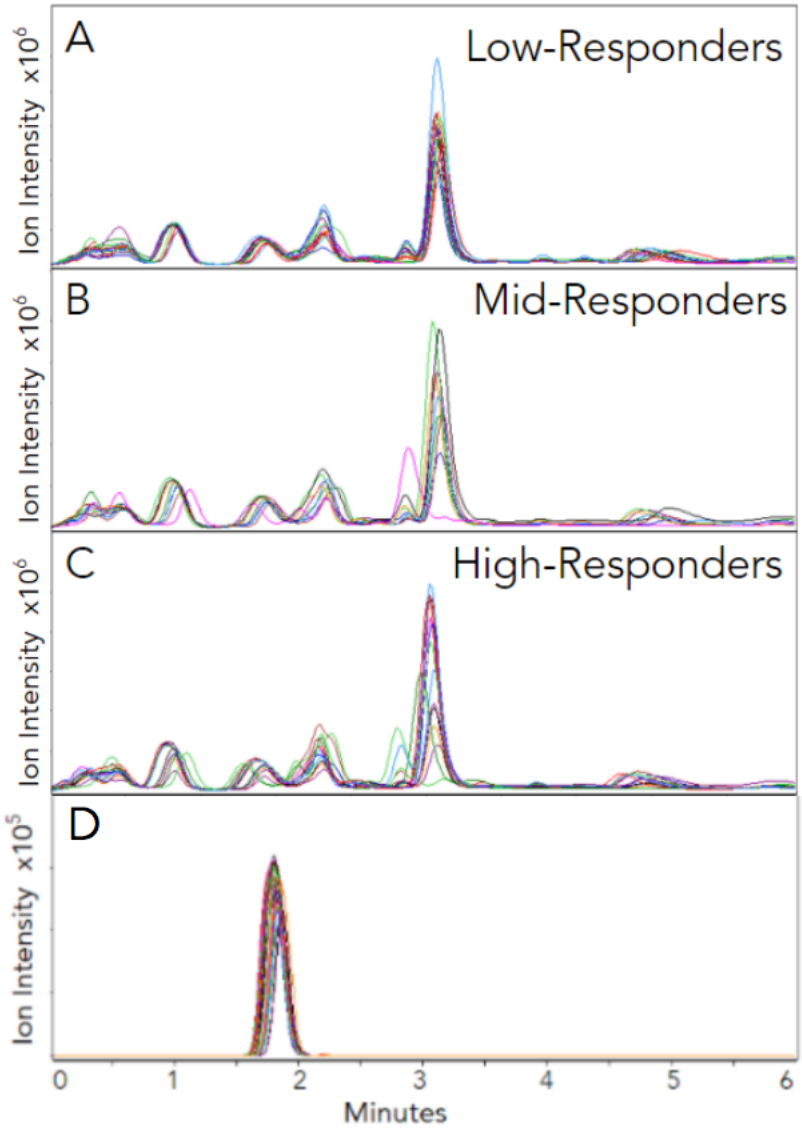
hydroxyhexyl) carbamic acid and the unknown metabolites with the m/z values of 240.151 and 82.026.

This analysis also shows how specific metabolites can be used to describe a dynamic inflammatory response. Previous studies describe the inverse relationship between creatine and cytokines such as IL1B (Bassit, et al., 2010). Our work takes this a step further in showing that the inverse relationship is most prominent in the first hour of an inflammatory response in individuals having a low-inflammation response. We propose that the increase in creatine is likely contributing to their lower cytokine response. Although the clinical significance of the magnitude of inflammatory responses is unknown, investigating the metabolic profiles of different responses can elucidate vital health information. A further examination could find parameters describing healthy magnitudes of inflammatory response for distinct groups and use this data to provide treatment.

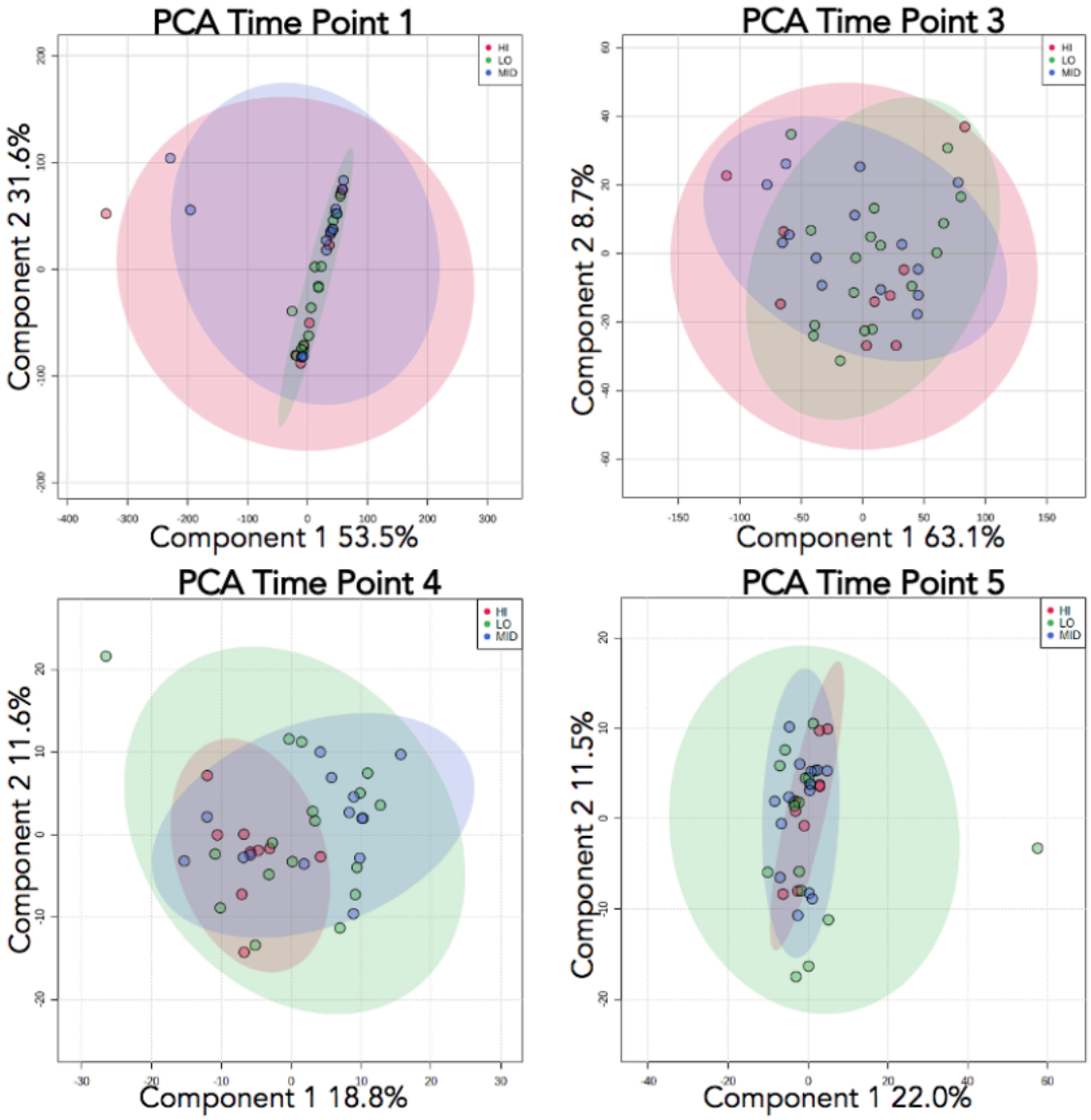
Inflammation responses are complex and variable between individuals so untangling the intricate web of interactions is challenging. Although many compounds in biological systems modulate or are modulated by immune responses, this analysis determined a suite of metabolites that may be predictive of immune response. This study also shows that temporal cytokine responses are correlated with temporal metabolic changes, which we refer to as a dynamic biosignature. By using a temporal interaction model, we were able to correlate changes in cytokines and metabolites which lead to distinct patterns of metabolic activity differentiating overweight and obese adults who responded differently to a dietary inflammation stimulus. In this way, a dynamic lens can provide insight into possible interventions for inflammatory diseases but can also pinpoint relevant interactions and time-dependent modalities.

Acknowledgements

The authors would like to acknowledge the support from a USDA-NIFA (2017-67018-26367) grant and a Montana State University Research Initiative (51040-MSUR12015-03) grant. We would also like to thank Dr. Ganesh Balasubramanian and Jesse Thomas from the Montana State University Proteomics, Metabolomics and Mass Spectrometry Facility for their mass spectrometry expertise and Dr. Sarah Bronsky for supervising sample collection. Funding for the Proteomics, Metabolomics and Mass Spectrometry Facility used int his publication was made possible in part by the MJ Murdock Charitable Trust and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM103474. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health.



Supplemental Figure 4.1. LCMS analysis of serum samples. Chromatograms for k-means separated inflammation groups at time point 2, corresponding to one-hour post meal. (A) Total ion chromatogram (TIC) for low-inflammation responders. (B) TIC for mid-inflammation responders. (C) TIC for high-inflammation responders. (D) Extracted ion chromatogram (EIC) for the m/z value of 114.065, corresponding to creatine (H+) for all participants in the study regardless of response grouping.



Supplemental Figure 4.2. Time point PCAs. 2D-PCA score plots for time points 1,3,4 and 5 for metabolomic profiles of the high-, mid- and low-response groups.

Supplemental Table 4.1. Participant characteristics

Participant Characteristics	LO n=17	MID n=14	HI n=9	p-value
Men/Women	8/9	5/9	3/6	NA
Age (years)	39.2±9.6	34.8±10.7	33.4±9.8	0.3
BMI (kg/m2)	30.8±2.6	30.1±1.3	30.2±1.6	0.67
Fat Mass (%)	35.7±7.6	35.8±6.7	36.8±6.6	0.92
VAT (l)	2.5±1.6	2.0±1.4	2.0±0.8	0.58
VO2 max (mL/kg/min)	42.4±9.6	44.6±8.3	40.9±11.9	0.67
HbA1c (%)	5.4±0.2	5.2±0.2	5.2±0.4	0.056
Fasting Cholesterol (mg/dL)	199.0±39.2	169.3±22.7	170.9±27.2	0.024
Fasting Triglycerides (mg/dL)	169.5±109.3	118.1±63.4	150.1±107.1	0.34
Fasting Glucose (mg/dL)	97.2±5.0	97.1±8.7	98.4±6.3	0.88

Supplemental Table 4.2. Retention time analysis for creatine

m/z	114.065
RT Mean	1.82 mins
RT Standard Deviation	0.02 mins
RT Coefficient of Variation	1.33

Supplemental Table 4.3. Metabolite coefficients of variation and ID matches

Metabolite ID	CV(%)	ID Match(%)
Acetamide	9.8	NA
C4H3NO	12.6	NA
Creatine	7.2	NA
Proline	4.7	NA
Organonitrogen compound C6H16N2O	17.1	96
Hydroxymethyluracil	8.9	53
Threitol	14.2	NA
4-Guanidinobutyrate	14.4	55
Guanine	16.9	NA
Carnitine	7.1	NA
Arylalkylamine compound C11H17N	6.5	96
(5-Amino-6-hydroxyhexyl)carbamate	14.7	45
Tiglylglycine	16.3	44
Butyryl Glycine	12.6	45
Azacyclic compound C10H17N5O2	9.2	90
Guanosine	18.3	NA
Histidinylhistidine	2.9	57

Literature Cited

AWWAD, H. M., GEISEL, J. & OBEID, R. 2016. Determination of trimethylamine, trimethylamine N-oxide, and taurine in human plasma and urine by UHPLC-MS/MS technique. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 1038**,** 12-18.

BAGHERI, M., FARZADFAR, F., QI, L., YEKANINEJAD, M. S., CHAMARI, M., ZELEZNIK, O. A., KALANTAR, Z., EBRAHIMI, Z., SHEIDAIE, A., KOLETZKO, B., UHL, O. & DJAZAYERY, A. 2018. Obesity-Related Metabolomic Profiles and Discrimination of Metabolically Unhealthy Obesity. *Journal of Proteome Research,* 17**,** 1452-1462.

BANSAL, S., BURING, J. E., RIFAI, N., MORA, S., SACKS, F. M. & RIDKER, P. M. 2007. Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. *Jama,* 298**,** 309-16.

BASSIT, R. A., CURI, R. & COSTA ROSA, L. F. 2008. Creatine supplementation reduces plasma levels of pro-inflammatory cytokines and PGE2 after a half-ironman competition. *Amino Acids,* 35**,** 425-31.

BAUMANN, H. & GAULDIE, J. 1994. The acute phase response. *Immunol Today,* 15**,** 74-80.

BOSY-WESTPHAL, A., SCHAUTZ, B., LATER, W., KEHAYIAS, J. J., GALLAGHER, D. & MÜLLER, M. J. 2013. What makes a BIA equation unique? Validity of eight-electrode multifrequency BIA to estimate body composition in a healthy adult population. *European Journal of Clinical Nutrition,* 67**,** S14-S21.

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. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology,* 30**,** 918-920.

CHANG, H. R. & BISTRIAN, B. 1998. The role of cytokines in the catabolic consequences of infection and injury. *JPEN J Parenter Enteral Nutr,* 22**,** 156-66.

CHONG, J., WISHART, D. S. & XIA, J. 2019. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics,* 68**,** e86.

COLGAN, S. P., CURTIS, V. F., LANIS, J. M. & GLOVER, L. E. 2015. Metabolic regulation of intestinal epithelial barrier during inflammation. *Tissue Barriers,* 3**,** e970936.

COLLETT, G., WOOD, A., ALEXANDER, M. Y., VARNUM, B. C., BOOT-HANDFORD, R. P., OHANIAN, V., OHANIAN, J., FRIDELL, Y. W. & CANFIELD, A. E. 2003. Receptor tyrosine kinase Axl modulates the osteogenic differentiation of pericytes. *Circ Res,* 92**,** 1123-9.

COLLETT, G. D. M. & CANFIELD, A. E. 2005. Angiogenesis and Pericytes in the Initiation of Ectopic Calcification. *Circulation Research,* 96**,** 930-938.

CONSIDINE, C. E. 2019. The Search for Clinically Useful Biomarkers of Complex Disease: A Data Analysis Perspective. *Metabolites,* 9.

CUPPS, T. R. & FAUCI, A. S. 1982. Corticosteroid-Mediated Immunoregulation in Man. *Immunological Reviews,* 65**,** 133-155.

DECARROZ, C., WAGNER, J. R., VAN LIER, J. E., KRISHNA, C. M., RIESZ, P. & CADET, J. 1986. Sensitized Photo-oxidation of Thymidine by 2-methyl-1,4-naphthoquinone. Characterization of the Stable Photoproducts. *International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine,* 50**,** 491-505.

DENIS, G. V. & OBIN, M. S. 2013. 'Metabolically healthy obesity': origins and implications. *Molecular aspects of medicine,* 34**,** 59-70.

DENVER, P. & MCCLEAN, P. L. 2018. Distinguishing normal brain aging from the development of Alzheimer's disease: inflammation, insulin signaling and cognition. *Neural regeneration research,* 13**,** 1719-1730.

DUDA, D. M., TU, C., FISHER, S. Z., AN, H., YOSHIOKA, C., GOVINDASAMY, L., LAIPIS, P. J., AGBANDJE-MCKENNA, M., SILVERMAN, D. N. & MCKENNA, R. 2005. Human carbonic anhydrase III: structural and kinetic study of catalysis and proton transfer. *Biochemistry,* 44**,** 10046-53.

DÜHRKOP, K., FLEISCHAUER, M., LUDWIG, M., AKSENOV, A. A., MELNIK, A. V., MEUSEL, M., DORRESTEIN, P. C., ROUSU, J. & BÖCKER, S. 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods,* 16**,** 299-302.

DÜHRKOP, K., SHEN, H., MEUSEL, M., ROUSU, J. & BÖCKER, S. 2015. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences,* 112**,** 12580.

ENGIN, A. 2017. The Pathogenesis of Obesity-Associated Adipose Tissue Inflammation. *In:* ENGIN, A. B. & ENGIN, A. (eds.) *Obesity and Lipotoxicity.* Cham: Springer International Publishing.

FITZPATRICK, M. & YOUNG, S. P. 2013. Metabolomics--a novel window into inflammatory disease. *Swiss medical weekly,* 143**,** w13743-w13743.

GOHLA, A. 2019. Do metabolic HAD phosphatases moonlight as protein phosphatases? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1866**,** 153-166.

GREGOR, M. F. & HOTAMISLIGIL, G. S. 2011. Inflammatory Mechanisms in Obesity. *Annual Review of Immunology,* 29**,** 415-445.

GRIGOROV, M. G. 2011. Analysis of Time Course Omics Datasets. *In:* MAYER, B. (ed.) *Bioinformatics for Omics Data: Methods and Protocols.* Totowa, NJ: Humana Press.

HALEY, M. J., MULLARD, G., HOLLYWOOD, K. A., COOPER, G. J., DUNN, W. B. & LAWRENCE, C. B. 2017. Adipose tissue and metabolic and inflammatory responses to stroke are altered in obese mice. *Dis Model Mech,* 10**,** 1229-1243.

HARTIGAN, J. A. & WONG, M. A. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 28**,** 100-108.

HERIEKA, M. & ERRIDGE, C. 2014. High-fat meal induced postprandial inflammation. *Mol Nutr Food Res,* 58**,** 136-46.

HOTAMISLIGIL, G. S. 2017. Inflammation, metaflammation and immunometabolic disorders. *Nature,* 542**,** 177.

KATAJAMAA, M., MIETTINEN, J. & OREŠIČ, M. 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics,* 22**,** 634-636.

KATAKAMI, N. 2018. Mechanism of Development of Atherosclerosis and Cardiovascular Disease in Diabetes Mellitus. *Journal of atherosclerosis and thrombosis,* 25**,** 27-39.

KAWAKAMI, T., KOIKE, A., MAEHARA, T., HAYASHI, T. & FUJIMORI, K. 2020. Bicarbonate enhances the inflammatory response by activating JAK/STAT signalling in LPS + IFN-γ-stimulated macrophages. *The Journal of Biochemistry,* 167**,** 623-631.

KERSTHOLT, M., VRIJMOETH, H., LACHMANDAS, E., OOSTING, M., LUPSE, M., FLONTA, M., DINARELLO, C. A., NETEA, M. G. & JOOSTEN, L. A. B. 2018. Role of glutathione metabolism in host defense against Borrelia burgdorferi infection. *Proceedings of the National Academy of Sciences,* 115**,** E2320.

LI, J. & WANG, C. 2019. GC/MS-based metabolomics strategy to analyze the effect of exercise intervention in diabetic rats. *Endocrine Connections,* 8**,** 654-660.

LISYOVÁ, J., CHANDOGA, J., JUNGOVÁ, P., REPISKÝ, M., KNAPKOVÁ, M., MACHKOVÁ, M., DLUHOLUCKÝ, S., BEHÚLOVÁ, D., ŠALIGOVÁ, J., POTOČŇÁKOVÁ, Ľ., LYSINOVÁ, M. & BÖHMER, D. 2018. An unusually high frequency of SCAD deficiency caused by two pathogenic variants in the ACADS gene and its relationship to the ethnic structure in Slovakia. *BMC medical genetics,* 19**,** 64-64.

MODRZEJEWSKA, M., GAWRONSKI, M., SKONIECZNA, M., ZARAKOWSKA, E., STARCZAK, M., FOKSINSKI, M., RZESZOWSKA-WOLNY, J., GACKOWSKI, D. & OLINSKI, R. 2016. Vitamin C enhances substantially formation of 5-hydroxymethyluracil in cellular DNA. *Free Radic Biol Med,* 101**,** 378-383.

NAMGALADZE, D. & BRÜNE, B. 2016. Macrophage fatty acid oxidation and its roles in macrophage polarization and fatty acid-induced inflammation. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids,* 1861**,** 1796-1807.

ORI, Y., ZINGERMAN, B., BERGMAN, M., BESSLER, H. & SALMAN, H. 2015. The effect of sodium bicarbonate on cytokine secretion in CKD patients with metabolic acidosis. *Biomed Pharmacother,* 71**,** 98-101.

PARK, C. H., EUN, C. S. & HAN, D. S. 2018. Intestinal microbiota, chronic inflammation, and colorectal cancer. *Intestinal research,* 16**,** 338-345.

POUPIN, N., TREMBLAY-FRANCO, M., AMIEL, A., CANLET, C., RÉMOND, D., DEBRAUWER, L., DARDEVET, D., THIELE, I., AURICH, M. K., JOURDAN, F., SAVARY-AUZELOUX, I. & POLAKOF, S. 2019. Arterio-venous metabolomics exploration reveals major changes across liver and intestine in the obese Yucatan minipig. *Scientific Reports,* 9**,** 12527.

REDDY, J. K. & SAMBASIVA RAO, M. 2006. Lipid Metabolism and Liver Inflammation. II. Fatty liver disease and fatty acid oxidation. *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 290**,** G852-G858.

RINGSEIS, R., EDER, K., MOOREN, F. & KRÜGER, K. 2015. Metabolic Signals and Innate Immune Activation in Obesity and Exercise. *Exercise immunology review,* 21**,** 2015.

ROZALSKI, R., GACKOWSKI, D., SIOMEK-GORECKA, A., STARCZAK, M., MODRZEJEWSKA, M., BANASZKIEWICZ, Z. & OLINSKI, R. 2015. Urinary 5-hydroxymethyluracil and 8-oxo-7,8-dihydroguanine as potential biomarkers in patients with colorectal cancer. *Biomarkers,* 20**,** 287-91.

SACCENTI, E., SMILDE, A. K. & CAMACHO, J. 2018. Group-wise ANOVA simultaneous component analysis for designed omics experiments. *Metabolomics,* 14**,** 73.

SAGE, A. P., TINTUT, Y. & DEMER, L. L. 2010. Regulatory mechanisms in vascular calcification. *Nature Reviews Cardiology,* 7**,** 528-536.

SATRIANO, J. 2004. Arginine pathways and the inflammatory response: interregulation of nitric oxide and polyamines: review article. *Amino Acids,* 26**,** 321-9.

SHARMA, A., TATE, M., MATHEW, G., VINCE, J. E., RITCHIE, R. H. & DE HAAN, J. B. 2018. Oxidative Stress and NLRP3-Inflammasome Activity as Significant Drivers of Diabetic Cardiovascular Complications: Therapeutic Implications. *Frontiers in physiology,* 9**,** 114-114.

SHINOHARA, Y. & TSUKIMOTO, M. 2018. Adenine Nucleotides Attenuate Murine T Cell Activation Induced by Concanavalin A or T Cell Receptor Stimulation. *Frontiers in Pharmacology,* 8.

SHIOMI, A. & USUI, T. 2015. Pivotal roles of GM-CSF in autoimmunity and inflammation. *Mediators Inflamm,* 2015**,** 568543.

SMILDE, A. K., JANSEN, J. J., HOEFSLOOT, H. C. J., LAMERS, R.-J. A. N., VAN DER GREEF, J. & TIMMERMAN, M. E. 2005. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics,* 21**,** 3043-3048.

SMITH, L. C. & NING, K. C. 1961. Effect of L-Histidine on Creatine, Histidine, and 1-Methylhistidine Excretion of Normal and Vit. E-Deficient Rabbits. *Proceedings of the Society for Experimental Biology and Medicine,* 107**,** 929-931.

SOGA, T., BARAN, R., SUEMATSU, M., UENO, Y., IKEDA, S., SAKURAKAWA, T., KAKAZU, Y., ISHIKAWA, T., ROBERT, M., NISHIOKA, T. & TOMITA, M. 2006. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *Journal of Biological Chemistry,* 281**,** 16768-16776.

STENVINKEL, P., HEIMBÜRGER, O., PAULTRE, F., DICZFALUSY, U., WANG, T., BERGLUND, L. & JOGESTRAND, T. 1999. Strong association between malnutrition,

inflammation, and atherosclerosis in chronic renal failure. *Kidney International,* 55**,** 1899-1911.

TOKMINA-LUKASZEWSKA, M., MOVAHED, N., LUSCZEK, E. R., MULIER, K. E., BEILMAN, G. J. & BOTHNER, B. 2014. Transformation of UPLC-MS Data Overcomes Extreme Variability in Urine Concentration and Metabolite Fold Change. *Current Metabolomics,* 2**,** 78-87.

TRICARICO, R., MADZO, J., SCHER, G., MAEGAWA, S., JELINEK, J., SCHER, C., CHANG, W.-C., NICOLAS, E., ZHOU, Y., SLIFKER, M., DEVARAJAN, K., CAI, K. Q., NAKAJIMA, P., XU, J., MANCUSO, P., DONEDDU, V., BAGELLA, L., INGRAM, J., BALACHANDRAN, S., PESHKOVA, I., KOLTSOVA, E., GRIVENNIKOV, S., YEN, T. J., ISSA, J.-P. & BELLACOSA, A. 2019. TET1 and TDG suppress intestinal tumorigenesis by down-regulating the inflammatory and immune response pathways. *bioRxiv***,** 676445.

WALLIMANN, T., TOKARSKA-SCHLATTNER, M. & SCHLATTNER, U. 2011. The creatine kinase system and pleiotropic effects of creatine. *Amino Acids,* 40**,** 1271-96.

WHEELOCK, C. E., GOSS, V. M., BALGOMA, D., NICHOLAS, B., BRANDSMA, J., SKIPP, P. J., SNOWDEN, S., BURG, D., AMICO, A., HORVATH, I., CHAIBOONCHOE, A., AHMED, H., BALLEREAU, S., ROSSIOS, C., CHUNG, K. F., MONTUSCHI, P., FOWLER, S. J., ADCOCK, I. M., POSTLE, A. D., DAHLÉN, S.-E., ROWE, A., STERK, P. J., AUFFRAY, C. & DJUKANOVIĆ, R. 2013. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *European Respiratory Journal,* 42**,** 802.

CHAPTER FIVE

CONCLUSIONS

This dissertation provides clear examples of the benefits that are offered by using temporal analysis of biological systems (**Figure 5.1**). The studies from chapters three and four would have had entirely different, and less dramatic results if only static interpretations had been used. It was the addition of time as a factor that revealed the important components of each puzzle. While not always possible, the benefits of temporal data are striking and should be included in standard systems biology methods.

By utilizing time as a variable, the importance of unique nitrogen and sulfur containing compounds was elucidated in thermal sediment. The unique geology of YNP combined with the effects of thermal vents creates novel hot springs environments. Creating a meta-analysis dataset indicated that microbes, especially archaea, correlate strongly with novel compounds. This analysis led to the hypothesis that archaea in these springs appear to have the ability to incorporate specific unique compounds in their metabolism. The loss of unique compounds then led to the loss of Archaea demonstrated in the population abundance and diversity data.

A shorter, but more intricate analysis of time-resolved serum data exhibited the importance of specific metabolites in the inflammation response in obese individuals. Compounds including creatine, hydroxymethyluracil and histidinylhistidine were able to provide mechanistic insight into the progression of inflammation. By creating models using response values and patterns, a dynamic biosignature was able to be created that can predict cytokine-mediated inflammatory responses. A cursory investigation of basal metabolites revealed no

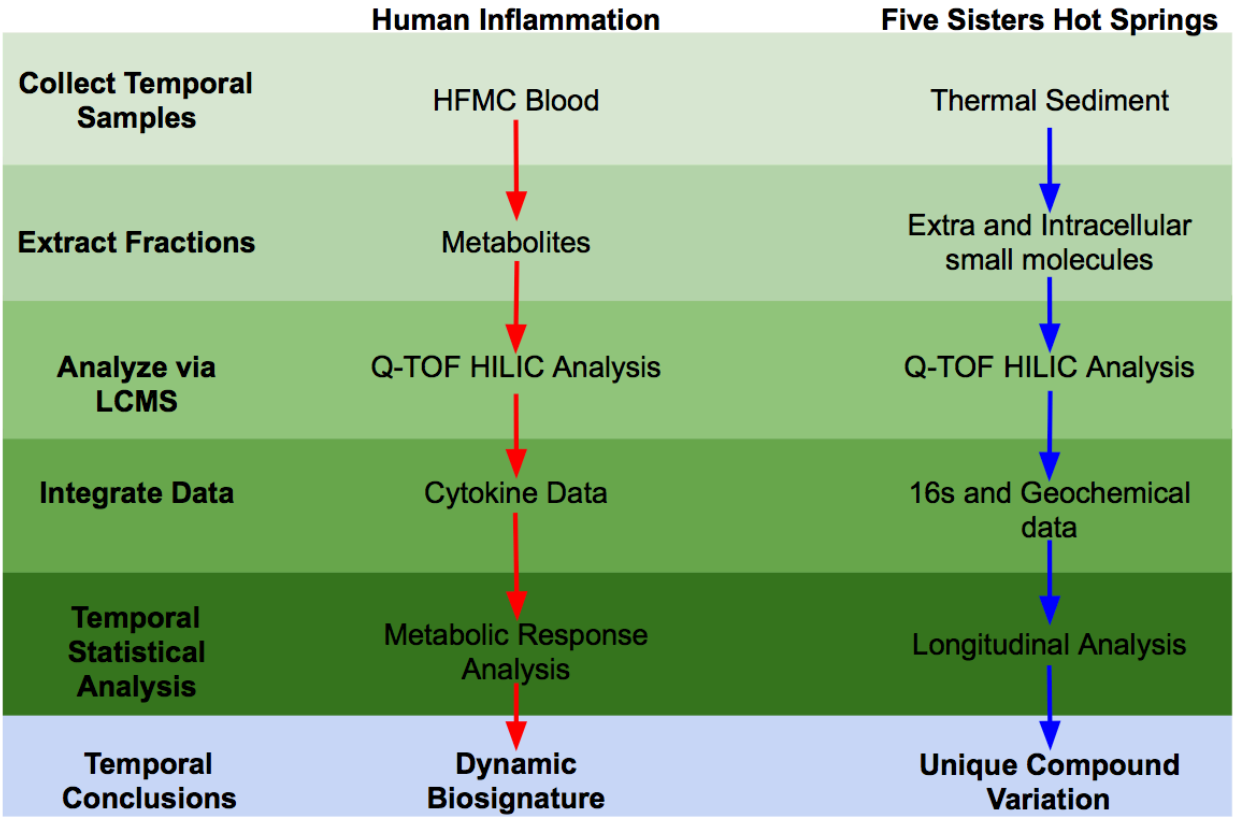


Figure 5.1. Schematic showing the steps taken in each study presented in this dissertation.

significant differences between the groups and analysis required the addition of a temporal component.

As exhibited in these studies, time-dependent data is often not only vital to systems biology research, it is scalable as well. The scalability of temporal systems biology has a greater range than described in this dissertation, varying from observing very small timescale metabolic rate changes in fluxomic analysis to determining genomic variations for phylogenetic analysis of species that lived thousands or millions of years apart (Canzler, Kapli). However, this scope of scalability coupled with multiple types of systems biology data and data formats has led to challenges integrating temporal aspects and in combining different omics datasets (Canzler, Veenstra).

Nevertheless, even with the current challenges, the incorporation of dynamism with the technological advances and computation power associated with systems biology is advancing (Veenstra). Dynamic interpretation revealed a new layer to each study in this dissertation which provided insights into biology. By relying solely on static systems biology data, important biological information can and will be missed or misinterpreted. The premise of systems biology is founded on a holistic principle of creating the most complete datasets for a class of biomolecules. Yet, this principle needs to be advanced to include a holistic view of temporal components and the dynamic nature of biology. As analysis of static systems biology becomes more intrinsic, innovation in resolving time dependent modalities can provide new perspective and through generation of novel methods to extrapolate temporal systems biology data, new frontiers can be explored.

Literature Cited

CANZLER, S., SCHOR, J., BUSCH, W., SCHUBERT, K., ROLLE-KAMPCZYK, U. E., SEITZ, H., KAMP, H., VON BERGEN, M., BUESEN, R. & HACKERMULLER, J. 2020. Prospects and challenges of multi-omics data integration in toxicology. *Archives of Toxicology,* 94**,** 371-388.

CASCANTE, M. & MARIN, S. 2008. Metabolomics and fluxomics approaches. *Essays Biochem,* 45**,** 67-81.

KAPLI, P., YANG, Z. & TELFORD, M. J. 2020. Phylogenetic tree building in the genomic age. *Nature Reviews Genetics,* 21**,** 428-444.

VEENSTRA, T. D. 2021. Omics in Systems Biology: Current Progress and Future Outlook. *PROTEOMICS,* 21**,** 2000235.

REFERENCES CITED

APPRILL, A., MCNALLY, S., PARSONS, R. & WEBER, L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology,* 75**,** 129-137.

AUTORE, G., CARUSO, A., MARZOCCO, S., NICOLAUS, B., PALLADINO, C., PINTO, A., POPOLO, A., SINICROPI, M. S., TOMMONARO, G. & SATURNINO, C. 2010. Acetamide derivatives with antioxidant activity and potential anti-inflammatory activity. *Molecules,* 15**,** 2028-38.

AWWAD, H. M., GEISEL, J. & OBEID, R. 2016. Determination of trimethylamine, trimethylamine N-oxide, and taurine in human plasma and urine by UHPLC-MS/MS technique. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 1038**,** 12-18.

BAGHERI, M., FARZADFAR, F., QI, L., YEKANINEJAD, M. S., CHAMARI, M., ZELEZNIK, O. A., KALANTAR, Z., EBRAHIMI, Z., SHEIDAIE, A., KOLETZKO, B., UHL, O. & DJAZAYERY, A. 2018. Obesity-Related Metabolomic Profiles and Discrimination of Metabolically Unhealthy Obesity. *Journal of Proteome Research,* 17**,** 1452-1462.

BANSAL, S., BURING, J. E., RIFAI, N., MORA, S., SACKS, F. M. & RIDKER, P. M. 2007. Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. *Jama,* 298**,** 309-16.

BARNS, S. M., FUNDYGA, R. E., JEFFRIES, M. W. & PACE, N. R. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 1609-1613.

BARSNES, H. & VAUDEL, M. 2018. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *J Proteome Res,* 17**,** 2552-2555.

BASSIT, R. A., CURI, R. & COSTA ROSA, L. F. 2008. Creatine supplementation reduces plasma levels of pro-inflammatory cytokines and PGE2 after a half-ironman competition. *Amino Acids,* 35**,** 425-31.

BAUMANN, H. & GAULDIE, J. 1994. The acute phase response. *Immunol Today,* 15**,** 74-80.

BAUMANN, S., KALKHOF, S., HACKERMULLER, J., OTTO, W., TOMM, J. M., WISSENBACH, D. K., ROLLE-KAMPCZYK, U. & VON BERGEN, M. 2013. Requirements and Perspectives for Integrating Metabolomics with other Omics Data. *Current Metabolomics,* 1**,** 15-27.

BOSY-WESTPHAL, A., SCHAUTZ, B., LATER, W., KEHAYIAS, J. J., GALLAGHER, D. & MÜLLER, M. J. 2013. What makes a BIA equation unique? Validity of eight-electrode multifrequency BIA to estimate body composition in a healthy adult population. *European Journal of Clinical Nutrition,* 67**,** S14-S21.

BROCK, T. D. 1967. Life at High Temperatures. *Science,* 158**,** 1012.

BUESCHER, J. M. & DRIGGERS, E. M. 2016. Integration of omics: more than the sum of its parts. *Cancer & metabolism,* 4**,** 4-4.

CANZLER, S., SCHOR, J., BUSCH, W., SCHUBERT, K., ROLLE-KAMPCZYK, U. E., SEITZ, H., KAMP, H., VON BERGEN, M., BUESEN, R. & HACKERMULLER, J. 2020. Prospects and challenges of multi-omics data integration in toxicology. *Archives of Toxicology,* 94**,** 371-388.

CAO, H. X., ZHANG, A. H., SUN, H., ZHOU, X. H., GUAN, Y., LIU, Q., KONG, L. & WANG, X. J. 2015. Metabolomics-proteomics profiles delineate metabolic changes in kidney fibrosis disease. *Proteomics,* 15**,** 3699-3710.

CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., LOZUPONE, C. A., TURNBAUGH, P. J., FIERER, N. & KNIGHT, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 4516-4522.

CASCANTE, M. & MARIN, S. 2008. Metabolomics and fluxomics approaches. *Essays Biochem,* 45**,** 67-81.

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. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology,* 30**,** 918-920.

CHANG, H. R. & BISTRIAN, B. 1998. The role of cytokines in the catabolic consequences of infection and injury. *JPEN J Parenter Enteral Nutr,* 22**,** 156-66.

CHONG, J., LIU, P., ZHOU, G. & XIA, J. 2020. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols,* 15**,** 799-821.

CHONG, J., WISHART, D. S. & XIA, J. 2019. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics,* 68**,** e86.

CHRISTIANSEN, R. L. 2001. The Quaternary and Pliocene Yellowstone Plateau volcanic field of Wyoming, Idaho, and Montana. *Professional Paper.* - ed.

CHUANG, H.-Y., HOFREE, M. & IDEKER, T. 2010. A decade of systems biology. *Annual review of cell and developmental biology,* 26**,** 721-744.

COLGAN, S. P., CURTIS, V. F., LANIS, J. M. & GLOVER, L. E. 2015. Metabolic regulation of intestinal epithelial barrier during inflammation. *Tissue Barriers,* 3**,** e970936.

COLLETT, G., WOOD, A., ALEXANDER, M. Y., VARNUM, B. C., BOOT-HANDFORD, R. P., OHANIAN, V., OHANIAN, J., FRIDELL, Y. W. & CANFIELD, A. E. 2003. Receptor tyrosine kinase Axl modulates the osteogenic differentiation of pericytes. *Circ Res,* 92**,** 1123-9.

COLLETT, G. D. M. & CANFIELD, A. E. 2005. Angiogenesis and Pericytes in the Initiation of Ectopic Calcification. *Circulation Research,* 96**,** 930-938.

CONSIDINE, C. E. 2019. The Search for Clinically Useful Biomarkers of Complex Disease: A Data Analysis Perspective. *Metabolites,* 9.

CUPPS, T. R. & FAUCI, A. S. 1982. Corticosteroid-Mediated Immunoregulation in Man. *Immunological Reviews,* 65**,** 133-155.

DECARROZ, C., WAGNER, J. R., VAN LIER, J. E., KRISHNA, C. M., RIESZ, P. & CADET, J. 1986. Sensitized Photo-oxidation of Thymidine by 2-methyl-1,4-naphthoquinone. Characterization of the Stable Photoproducts. *International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine,* 50**,** 491-505.

DECASTRO, M.-E., RODRÍGUEZ-BELMONTE, E. & GONZÁLEZ-SISO, M.-I. 2016a. Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Frontiers in microbiology,* 7**,** 1521-1521.

DECASTRO, M. E., RODRIGUEZ-BELMONTE, E. & GONZALEZ-SISO, M. I. 2016b. Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Frontiers in Microbiology,* 7**,** 21.

DENIS, G. V. & OBIN, M. S. 2013. 'Metabolically healthy obesity': origins and implications. *Molecular aspects of medicine,* 34**,** 59-70.

DENVER, P. & MCCLEAN, P. L. 2018. Distinguishing normal brain aging from the development of Alzheimer's disease: inflammation, insulin signaling and cognition. *Neural regeneration research,* 13**,** 1719-1730.

DETTMER, K., ARONOV, P. A. & HAMMOCK, B. D. 2007a. Mass spectrometry-based metabolomics. *Mass Spectrom Rev,* 26**,** 51-78.

DITTMAR, T., KOCH, B., HERTKORN, N. & KATTNER, G. 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography-Methods,* 6**,** 230-235.

DOMON, B. & AEBERSOLD, R. 2006. Review - Mass spectrometry and protein analysis. *Science,* 312**,** 212-217.

DUDA, D. M., TU, C., FISHER, S. Z., AN, H., YOSHIOKA, C., GOVINDASAMY, L., LAIPIS, P. J., AGBANDJE-MCKENNA, M., SILVERMAN, D. N. & MCKENNA, R. 2005. Human carbonic anhydrase III: structural and kinetic study of catalysis and proton transfer. *Biochemistry,* 44**,** 10046-53.

DÜHRKOP, K., FLEISCHAUER, M., LUDWIG, M., AKSENOV, A. A., MELNIK, A. V., MEUSEL, M., DORRESTEIN, P. C., ROUSU, J. & BÖCKER, S. 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods,* 16**,** 299-302.

DÜHRKOP, K., SHEN, H., MEUSEL, M., ROUSU, J. & BÖCKER, S. 2015. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences,* 112**,** 12580.

EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* 26**,** 2460-2461.

EDGAR, R. C. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv,* Preprint.

ENGIN, A. 2017. The Pathogenesis of Obesity-Associated Adipose Tissue Inflammation. *In:* ENGIN, A. B. & ENGIN, A. (eds.) *Obesity and Lipotoxicity.* Cham: Springer International Publishing.

FARAJI, M., YAMINI, Y. & GHOLAMI, M. 2019. Recent Advances and Trends in Applications of Solid-Phase Extraction Techniques in Food and Environmental Analysis. *Chromatographia,* 82**,** 1207-1249.

FENG, X. J., LIU, X., LUO, Q. M. & LIU, B. F. 2008. MASS SPECTROMETRY IN SYSTEMS BIOLOGY: AN OVERVIEW. *Mass Spectrometry Reviews,* 27**,** 635-660.

FITZPATRICK, M. & YOUNG, S. P. 2013. Metabolomics--a novel window into inflammatory disease. *Swiss medical weekly,* 143**,** w13743-w13743.

FOURNIER, R., THOMPSON, MICHAEL J., HUTCHINSON, RODERICK A. 1992. The geochemistry of hot spring waters at Norris Geyser Basin, Yellowstone National Park. *International symposium on water-rock interactions.*

GASPAR, A., HARIR, M., HERTKORN, N. & SCHMITT-KOPPLIN, P. 2010. Preparative free-flow electrophoretic offline ESI-Fourier transform ion cyclotron resonance/MS analysis of Suwannee River fulvic acid. *Electrophoresis,* 31**,** 2070-2079.

GIBSON, M. L. & HINMAN, N. W. 2013. Mixing of hydrothermal water and groundwater near hot springs, Yellowstone National Park (USA): hydrology and geochemistry. *Hydrogeology Journal,* 21**,** 919-933.

GOHLA, A. 2019. Do metabolic HAD phosphatases moonlight as protein phosphatases? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1866**,** 153-166.

GONSIOR, M., HERTKORN, N., HINMAN, N., DVORSKI, S. E. M., HARIR, M., COOPER, W. J. & SCHMITT-KOPPLIN, P. 2018. Yellowstone Hot Springs are Organic Chemodiversity Hot Spots. *Scientific Reports,* 8**,** 14155.

GREGOR, M. F. & HOTAMISLIGIL, G. S. 2011. Inflammatory Mechanisms in Obesity. *Annual Review of Immunology,* 29**,** 415-445.

GRIGOROV, M. G. 2011. Analysis of Time Course Omics Datasets. *In:* MAYER, B. (ed.) *Bioinformatics for Omics Data: Methods and Protocols.* Totowa, NJ: Humana Press.

GROMIHA, M. M. & SURESH, M. X. 2008. Discrimination of mesophilic and thermophilic proteins using machine learning algorithms. *Proteins,* 70**,** 1274-9.

HALEY, M. J., MULLARD, G., HOLLYWOOD, K. A., COOPER, G. J., DUNN, W. B. & LAWRENCE, C. B. 2017. Adipose tissue and metabolic and inflammatory responses to stroke are altered in obese mice. *Dis Model Mech,* 10**,** 1229-1243.

HARTIGAN, J. A. & WONG, M. A. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 28**,** 100-108.

HEDLUND, B. P., MURUGAPIRAN, S. K., ALBA, T. W., LEVY, A., DODSWORTH, J. A., GOERTZ, G. B., IVANOVA, N. & WOYKE, T. 2015. Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. *Current Opinion in Microbiology,* 25**,** 136-145.

HERIEKA, M. & ERRIDGE, C. 2014. High-fat meal induced postprandial inflammation. *Mol Nutr Food Res,* 58**,** 136-46.

HOOD, L. 2003. Systems biology: integrating technology, biology, and computation. *Mechanisms of Ageing and Development,* 124**,** 9-16.

HOTAMISLIGIL, G. S. 2017. Inflammation, metaflammation and immunometabolic disorders. *Nature,* 542**,** 177.

HUBER, W., CAREY, V. J., GENTLEMAN, R., ANDERS, S., CARLSON, M., CARVALHO, B. S., BRAVO, H. C., DAVIS, S., GATTO, L., GIRKE, T., GOTTARDO, R., HAHNE, F., HANSEN, K. D., IRIZARRY, R. A., LAWRENCE, M., LOVE, M. I., MACDONALD, J., OBENCHAIN, V., OLEŚ, A. K., PAGÈS, H., REYES, A., SHANNON, P., SMYTH, G. K., TENENBAUM, D., WALDRON, L. & MORGAN, M. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods,* 12**,** 115-121.

IDEKER, T., GALITSKI, T. & HOOD, L. 2001. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet,* 2**,** 343-72.

JING, L. & CHENGJI, W. 2019. GC/MS-based metabolomics strategy to analyze the effect of exercise intervention in diabetic rats. *Endocr Connect,* 8**,** 654-660.

KAPLI, P., YANG, Z. & TELFORD, M. J. 2020. Phylogenetic tree building in the genomic age. *Nature Reviews Genetics,* 21**,** 428-444.

KATAJAMAA, M., MIETTINEN, J. & OREŠIČ, M. 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics,* 22**,** 634-636.

KATAKAMI, N. 2018. Mechanism of Development of Atherosclerosis and Cardiovascular Disease in Diabetes Mellitus. *Journal of atherosclerosis and thrombosis,* 25**,** 27-39.

KAWAKAMI, T., KOIKE, A., MAEHARA, T., HAYASHI, T. & FUJIMORI, K. 2020. Bicarbonate enhances the inflammatory response by activating JAK/STAT signalling in LPS + IFN-γ-stimulated macrophages. *The Journal of Biochemistry,* 167**,** 623-631.

KERSTHOLT, M., VRIJMOETH, H., LACHMANDAS, E., OOSTING, M., LUPSE, M., FLONTA, M., DINARELLO, C. A., NETEA, M. G. & JOOSTEN, L. A. B. 2018. Role

of glutathione metabolism in host defense against Borrelia burgdorferi infection. *Proceedings of the National Academy of Sciences,* 115**,** E2320.

KEW, W., BLACKBURN, J. W. T., CLARKE, D. J. & UHRÍN, D. 2017. Interactive van Krevelen diagrams – Advanced visualisation of mass spectrometry data of complex mixtures. *Rapid Communications in Mass Spectrometry,* 31**,** 658-662.

KIM, S., KRAMER, R. W. & HATCHER, P. G. 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Analytical Chemistry,* 75**,** 5336-5344.

LEI, Z. T., HUHMAN, D. V. & SUMNER, L. W. 2011. Mass Spectrometry Strategies in Metabolomics. *Journal of Biological Chemistry,* 286**,** 25435-25442.

LI, J. & WANG, C. 2019. GC/MS-based metabolomics strategy to analyze the effect of exercise intervention in diabetic rats. *Endocrine Connections,* 8**,** 654-660.

LISYOVÁ, J., CHANDOGA, J., JUNGOVÁ, P., REPISKÝ, M., KNAPKOVÁ, M., MACHKOVÁ, M., DLUHOLUCKÝ, S., BEHÚLOVÁ, D., ŠALIGOVÁ, J., POTOČŇÁKOVÁ, Ľ., LYSINOVÁ, M. & BÖHMER, D. 2018. An unusually high frequency of SCAD deficiency caused by two pathogenic variants in the ACADS gene and its relationship to the ethnic structure in Slovakia. *BMC medical genetics,* 19**,** 64-64.

LÓPEZ-LÓPEZ, O., CERDÁN, M.-E. & GONZÁLEZ-SISO, M.-I. 2015. Thermus thermophilus as a Source of Thermostable Lipolytic Enzymes. *Microorganisms,* 3**,** 792-808.

MODRZEJEWSKA, M., GAWRONSKI, M., SKONIECZNA, M., ZARAKOWSKA, E., STARCZAK, M., FOKSINSKI, M., RZESZOWSKA-WOLNY, J., GACKOWSKI, D. & OLINSKI, R. 2016. Vitamin C enhances substantially formation of 5-hydroxymethyluracil in cellular DNA. *Free Radic Biol Med,* 101**,** 378-383.

MUELLER, R. C., PEACH, J. T., SKORUPA, D. J., COPIÉ, V., BOTHNER, B. & PEYTON, B. M. 2020. An emerging view of the diversity, ecology, and function of Archaea in alkaline hydrothermal environments. *FEMS Microbiology Ecology*.

NAMGALADZE, D. & BRÜNE, B. 2016. Macrophage fatty acid oxidation and its roles in macrophage polarization and fatty acid-induced inflammation. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids,* 1861**,** 1796-1807.

OFFRE, P., SPANG, A. & SCHLEPER, C. 2013. Archaea in biogeochemical cycles. *Annu Rev Microbiol,* 67**,** 437-57.

ORI, Y., ZINGERMAN, B., BERGMAN, M., BESSLER, H. & SALMAN, H. 2015. The effect of sodium bicarbonate on cytokine secretion in CKD patients with metabolic acidosis. *Biomed Pharmacother,* 71**,** 98-101.

PALA, C., MOLARI, M., NIZZOLI, D., BARTOLI, M., VIAROLI, P. & MANINI, E. 2018. Environmental Drivers Controlling Bacterial and Archaeal Abundance in the Sediments of a Mediterranean Lagoon Ecosystem. *Current Microbiology,* 75**,** 1147-1155.

PARADA, A. E., NEEDHAM, D. M. & FUHRMAN, J. A. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology,* 18**,** 1403-1414.

PARK, C. H., EUN, C. S. & HAN, D. S. 2018. Intestinal microbiota, chronic inflammation, and colorectal cancer. *Intestinal research,* 16**,** 338-345.

PATEL, A. K., SINGHANIA, R. R., SIM, S. J. & PANDEY, A. 2019. Thermostable cellulases: Current status and perspectives. *Bioresour Technol,* 279**,** 385-392.

PATTI, G. J., YANES, O. & SIUZDAK, G. 2012. Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology,* 13**,** 263-269.

PLUSKAL, T., CASTILLO, S., VILLAR-BRIONES, A. & ORESIC, M. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics,* 11**,** 395.

POUPIN, N., TREMBLAY-FRANCO, M., AMIEL, A., CANLET, C., RÉMOND, D., DEBRAUWER, L., DARDEVET, D., THIELE, I., AURICH, M. K., JOURDAN, F., SAVARY-AUZELOUX, I. & POLAKOF, S. 2019. Arterio-venous metabolomics exploration reveals major changes across liver and intestine in the obese Yucatan minipig. *Scientific Reports,* 9**,** 12527.

RAFTERY, D. 2016. *Mass Spectrometry in Metabolomics*, Springer.

RAWLE, R. A., HAMERLY, T., TRIPET, B. P., GIANNONE, R. J., WURCH, L., HETTICH, R. L., PODAR, M., COPIÉ, V. & BOTHNER, B. 2017. Multi-omics analysis provides insight to the Ignicoccus hospitalis-Nanoarchaeum equitans association. *Biochimica et biophysica acta. General subjects,* 1861**,** 2218-2227.

REDDY, J. K. & SAMBASIVA RAO, M. 2006. Lipid Metabolism and Liver Inflammation. II. Fatty liver disease and fatty acid oxidation. *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 290**,** G852-G858.

REN, J. L., ZHANG, A. H., KONG, L. & WANG, X. J. 2018. Advances in mass spectrometry-based metabolomics for investigation of metabolites. *Rsc Advances,* 8**,** 22335-22350.

RICHELLE, A., DAVID, B., DEMAEGD, D., DEWERCHIN, M., KINET, R., MORREALE, A., PORTELA, R., ZUNE, Q. & VON STOSCH, M. 2020. Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. *npj Systems Biology and Applications,* 6**,** 6.

RINGSEIS, R., EDER, K., MOOREN, F. & KRÜGER, K. 2015. Metabolic Signals and Innate Immune Activation in Obesity and Exercise. *Exercise immunology review,* 21**,** 2015.

ROHART, F., GAUTIER, B., SINGH, A. & LÊ CAO, K.-A. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology,* 13**,** e1005752-e1005752.

ROWE, J. J., FOURNIER, R. & MOREY, G. Chemical analysis of thermal waters in Yellowstone National Park, Wyoming, 1960-65. 1973.

ROZALSKI, R., GACKOWSKI, D., SIOMEK-GORECKA, A., STARCZAK, M., MODRZEJEWSKA, M., BANASZKIEWICZ, Z. & OLINSKI, R. 2015. Urinary 5-hydroxymethyluracil and 8-oxo-7,8-dihydroguanine as potential biomarkers in patients with colorectal cancer. *Biomarkers,* 20**,** 287-91.

SACCENTI, E., SMILDE, A. K. & CAMACHO, J. 2018. Group-wise ANOVA simultaneous component analysis for designed omics experiments. *Metabolomics,* 14**,** 73.

SAGE, A. P., TINTUT, Y. & DEMER, L. L. 2010. Regulatory mechanisms in vascular calcification. *Nature Reviews Cardiology,* 7**,** 528-536.

SAHAY, H., YADAV, A. N., SINGH, A. K., SINGH, S., KAUSHIK, R. & SAXENA, A. K. 2017. Hot springs of Indian Himalayas: potential sources of microbial diversity and thermostable hydrolytic enzymes. *3 Biotech,* 7**,** 118.

SANGUINETTI, M. & POSTERARO, B. 2016. Mass spectrometry applications in microbiology beyond microbe identification: progress and potential. *Expert Review of Proteomics,* 13**,** 965-977.

SATRIANO, J. 2004. Arginine pathways and the inflammatory response: interregulation of nitric oxide and polyamines: review article. *Amino Acids,* 26**,** 321-9.

SHARMA, A., TATE, M., MATHEW, G., VINCE, J. E., RITCHIE, R. H. & DE HAAN, J. B. 2018. Oxidative Stress and NLRP3-Inflammasome Activity as Significant Drivers of Diabetic Cardiovascular Complications: Therapeutic Implications. *Frontiers in physiology,* 9**,** 114-114.

SHINOHARA, Y. & TSUKIMOTO, M. 2018. Adenine Nucleotides Attenuate Murine T Cell Activation Induced by Concanavalin A or T Cell Receptor Stimulation. *Frontiers in Pharmacology,* 8.

SHIOMI, A. & USUI, T. 2015. Pivotal roles of GM-CSF in autoimmunity and inflammation. *Mediators Inflamm,* 2015**,** 568543.

SMILDE, A. K., JANSEN, J. J., HOEFSLOOT, H. C. J., LAMERS, R.-J. A. N., VAN DER GREEF, J. & TIMMERMAN, M. E. 2005. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics,* 21**,** 3043-3048.

SMITH, L. C. & NING, K. C. 1961. Effect of L-Histidine on Creatine, Histidine, and 1-Methylhistidine Excretion of Normal and Vit. E-Deficient Rabbits. *Proceedings of the Society for Experimental Biology and Medicine,* 107**,** 929-931.

SMITH, H. J., TIGGES, M., D'ANDRILLI, J., PARKER, A., BOTHNER, B. & FOREMAN, C. M. 2018. Dynamic processing of DOM: Insight from exometabolomics, fluorescence spectroscopy, and mass spectrometry. *Limnology and Oceanography Letters,* 3**,** 225-235.

SOGA, T., BARAN, R., SUEMATSU, M., UENO, Y., IKEDA, S., SAKURAKAWA, T., KAKAZU, Y., ISHIKAWA, T., ROBERT, M., NISHIOKA, T. & TOMITA, M. 2006. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *Journal of Biological Chemistry,* 281**,** 16768-16776.

SPENCER, R. G. M., GUO, W., RAYMOND, P. A., DITTMAR, T., HOOD, E., FELLMAN, J. & STUBBINS, A. 2014. Source and biolability of ancient dissolved organic matter in glacier and lake ecosystems on the Tibetan Plateau. *Geochimica et Cosmochimica Acta,* 142**,** 64-74.

STENVINKEL, P., HEIMBÜRGER, O., PAULTRE, F., DICZFALUSY, U., WANG, T., BERGLUND, L. & JOGESTRAND, T. 1999. Strong association between malnutrition, inflammation, and atherosclerosis in chronic renal failure. *Kidney International,* 55**,** 1899-1911.

THE UNIPROT, C. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research,* 49**,** D480-D489.

THIEL, V., WOOD, J. M., OLSEN, M. T., TANK, M., KLATT, C. G., WARD, D. M. & BRYANT, D. A. 2016. The Dark Side of the Mushroom Spring Microbial Mat: Life in the Shadow of Chlorophototrophs. I. Microbial Diversity Based on 16S rRNA Gene Amplicons and Metagenomic Sequencing. *Frontiers in Microbiology,* 7.

TOKMINA-LUKASZEWSKA, M., MOVAHED, N., LUSCZEK, E. R., MULIER, K. E., BEILMAN, G. J. & BOTHNER, B. 2014. Transformation of UPLC-MS Data Overcomes Extreme Variability in Urine Concentration and Metabolite Fold Change. *Current Metabolomics,* 2**,** 78-87.

TRICARICO, R., MADZO, J., SCHER, G., MAEGAWA, S., JELINEK, J., SCHER, C., CHANG, W.-C., NICOLAS, E., ZHOU, Y., SLIFKER, M., DEVARAJAN, K., CAI, K. Q., NAKAJIMA, P., XU, J., MANCUSO, P., DONEDDU, V., BAGELLA, L., INGRAM, J., BALACHANDRAN, S., PESHKOVA, I., KOLTSOVA, E., GRIVENNIKOV, S., YEN, T. J., ISSA, J.-P. & BELLACOSA, A. 2019. TET1 and TDG suppress intestinal tumorigenesis by down-regulating the inflammatory and immune response pathways. *bioRxiv***,** 676445.

VAN DER MEER, M. T. J., SCHOUTEN, S., BATESON, M. M., NÜBEL, U., WIELAND, A., KÜHL, M., DE LEEUW, J. W., SINNINGHE DAMSTÉ, J. S. & WARD, D. M. 2005. Diel Variations in Carbon Metabolism by Green Nonsulfur-Like Bacteria in Alkaline Siliceous Hot Spring Microbial Mats from Yellowstone National Park. *Applied and Environmental Microbiology,* 71**,** 3978.

VAUDEL, M., BURKHART, J. M., ZAHEDI, R. P., OVELAND, E., BERVEN, F. S., SICKMANN, A., MARTENS, L. & BARSNES, H. 2015. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nature Biotechnology,* 33**,** 22-24.

VEENSTRA, T. D. 2021. Omics in Systems Biology: Current Progress and Future Outlook. *PROTEOMICS,* 21**,** 2000235.

VERMA, P., AJAR NATH YADAV, LIVLEEN SHUKLA, ANIL KUMAR SAXENA AND ARCHNA SUMAN 2015. Hydrolytic enzymes production by thermotolerant Bacillus altitudinis IARI-MB-9 and Gulbenkiania mobilis IARI-MB-18 isolated from Manikaran hot springs. *International Journal of Advanced Research,* 3**,** 1241-1250.

VILLA, A. & SONIS, S. T. 2020. 2 - System biology. *In:* SONIS, S. T. & VILLA, A. (eds.) *Translational Systems Medicine and Oral Disease.* Academic Press.

WALLIMANN, T., TOKARSKA-SCHLATTNER, M. & SCHLATTNER, U. 2011. The creatine kinase system and pleiotropic effects of creatine. *Amino Acids,* 40**,** 1271-96.

WANG, K., ZHANG, Y., HUANG, R.-J., CAO, J. & HOFFMANN, T. 2018. UHPLC-Orbitrap mass spectrometric characterization of organic aerosol from a central European city (Mainz, Germany) and a Chinese megacity (Beijing). *Atmospheric Environment,* 189**,** 22-29.

WANG, X., HAYECK, N., BRÜGGEMANN, M., YAO, L., CHEN, H., ZHANG, C., EMMELIN, C., CHEN, J., GEORGE, C. & WANG, L. 2017. Chemical Characteristics of Organic Aerosols in Shanghai: A Study by Ultrahigh-Performance Liquid

Chromatography Coupled With Orbitrap Mass Spectrometry. *Journal of Geophysical Research: Atmospheres,* 122**,** 11,703-11,722.

WHEELOCK, C. E., GOSS, V. M., BALGOMA, D., NICHOLAS, B., BRANDSMA, J., SKIPP, P. J., SNOWDEN, S., BURG, D., AMICO, A., HORVATH, I., CHAIBOONCHOE, A., AHMED, H., BALLEREAU, S., ROSSIOS, C., CHUNG, K. F., MONTUSCHI, P., FOWLER, S. J., ADCOCK, I. M., POSTLE, A. D., DAHLÉN, S.-E., ROWE, A., STERK, P. J., AUFFRAY, C. & DJUKANOVIĆ, R. 2013. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *European Respiratory Journal,* 42**,** 802.

WINGEN, M., JAEGER, K. E., GENSCH, T. & DREPPER, T. 2017. Novel Thermostable Flavin-binding Fluorescent Proteins from Thermophilic Organisms. *Photochem Photobiol,* 93**,** 849-856.

ZUPANIC, A., BERNSTEIN, H. C. & HEILAND, I. 2020. Systems biology: current status and challenges. *Cellular and Molecular Life Sciences,* Feb**,** 379-380.