UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

A multi-cultural characterization of human-environmental interaction through metabolomic and microbiome profiling

Permalink

https://escholarship.org/uc/item/4t04z4rp

Author

Kapono, Clifford

Publication Date

2018

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

A multi-cultural characterization of human-environmental interaction through metabolomic and microbiome profiling

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Chemistry

by

Clifford A. Kapono

Committee in charge:

Professor Pieter Dorrestein, Chair
Professor Kamil Godula
Professor Elizabeth Komives
Professor Bradley Moore
Professor Jennifer Smith
Professor Yitzhak Tor

2018

Copyright

Clifford A. Kapono, 2018

All rights reserved.

The Dissertation of Clifford A. Kapono is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego
2018

DEDICATION

No ka lahui ‘o Hawai‘i. E ola. E ola mau.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor, Pieter Dorrestein, who has encouraged, supported, and inspired me throughout my graduate career and has facilitated a collaborative and exploratory laboratory environment that I have greatly enjoyed. I would also like to thank my committee members Bradley Moore, Elizabeth Komives, Yitzhak Tor, Kamil Godula and Jennifer Smith for believing in my commitment to science.

I would like to thank Jeff Rances from the chemistry department for his continual support throughout my time at UCSD. I would like to express my appreciation to the many labs at SIO that have allowed me to perform research in their lab such as the Gerwick lab, Sandin lab, Smith lab, Moore lab, and Azam lab. Individuals whom I owe much of my academic progress include Paul Boudreau, Matt Bertin, Samantha Mascuch, Bill Gerwick, Lena Gerwick, Clint Edwards, Yoan Eynaud, Brian Zgliczynski, Kate Furby, Stu Sandin, Jia Jia Zhang, Vinny Agarwal, Ryan Guillmette, and Farooq Azam. Additionally, the research presented would now be possible without positive and dedicated support of my fellow lab members in the Dorrestein lab like Mike Meehan, Alexey Melnik, Amina Bouslimani, Laura Sanchez, Vanessa Phelan, Don Nguyen, Fernando Vargas, Neha Garg, Ming Wang, Tommy Boy Hoffmann, Louis-Felix Nothias, Ricardo da Silva, Alexander Aksenov, Andrés Mauricio Caraballo Rodríguez, Tal Luzzatto Knaan, Jane Yang, and Theodore Alexandrov. Lastly, I would like to thank my family and friends who have supported me throughout my entire journey.

Chapter 2, in full, is a reprint of materials as it appears in “Creating a 3D microbial and chemical snapshot of a human habitat” in *Scientific Reports*, 2018, Kapono, C. A., Morton, J. T., Bouslimani, A., Melnik, A. V., Orlinsky, K., Knaan, T. L, Garg, N., Vazquez-Baeza, Y., Protsyuk, I., Janssen, S., Zhu, Q., Alexandrov, T., Smarr, L., Knight, R., Dorrestein, P.C.. The dissertation author was the primary investigator and author of this manuscript.

Chapter 3 is currently being prepared for submission of the material. Kapono, C.A., Burns, J.H.R., Ackermann, G., Humphrey, G., Takabayashi, M., Knight, R., Dorrestein, P.C. The dissertation author was the primary investigator and author of this manuscript.

Chapter 4 is currently being prepared for submission of the material. Kapono, C.A., Tripathi, A., Callewaert, C., Mcdonald, D., Sanders, J., Da Silva, R., Nothias, L.F., Ernst, M., Petras, D., Deright Goldasich, L.D., Akermann, G., Knight, R., Dorrestein, P.C. The dissertation author was the primary investigator and author of this manuscript.

The work presented in this dissertation was funded by the Tribal Initiative Fellowship, Programs of Interdisciplinary Environmental Research Fellowship and the UCSD Global Health Institute Graduate Research Fellowship.

VITA

2009 University of Hawai‘i at Manoa, Honolulu, Hawai‘i, US

B.S in Plant and Environmental Biotechnology

Research Advisor: Prof. Sean S. Callahan

2012 University of Hawai‘i at Manoa, Honolulu, Hawai‘i, US

M.S in Molecular Bioscience and Bioengineering

Research Advisor: Prof. Jon-Paul Bingham

2018 University of California, San Diego, La Jolla, CA

Ph.D. in Chemistry

Research Advisor: Prof. Pieter C. Dorrestein

PUBLICATIONS

1. **Kapono CA**, Morton JT, Bouslimani A, Melnik AV, Orlinsky K, Knaan TL, Garg N, Vázquez-Baeza Y, Protsyuk I, Janssen S, Zhu Q. Creating a 3D microbial and chemical snapshot of a human habitat. Scientific reports. **2018** Feb 27;8(1):3669.

2. Floros DJ, Petras D, **Kapono CA**, Melnik AV, Ling TJ, Knight R, Dorrestein PC. Mass spectrometry based molecular 3D-cartography of plant metabolites. Frontiers in plant science. **2017** Mar 29;8:429.

3. Petras D, Nothias LF, Quinn RA, Alexandrov T, Bandeira N, Bouslimani A, Castro-Falcón G, Chen L, Dang T, Floros DJ, Hook V., **Kapono CA.** Mass spectrometry-based visualization of molecules associated with human habitats. Analytical chemistry. **2016** Oct 24;88(22):10775-84.

4. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, **Kapono CA**, Luzzatto-Knaan T, Porto C. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nature biotechnology. **2016** Aug;34(8):828.

5. Garg N, **Kapono CA**, Lim YW, Koyama N, Vermeij MJ, Conrad D, Rohwer F, Dorrestein PC. Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures. International journal of mass spectrometry. **2015** Feb 1;377:719-27.

6. Sandoval-Calderón M, Nguyen DD, **Kapono CA**, Herron P, Dorrestein PC, Sohlenkamp C. Plasticity of Streptomyces coelicolor membrane composition under different growth conditions and during development. Frontiers in microbiology. **2015** Dec 22;6:1465.

7. **Kapono CA**, Thapa P, Cabalteja CC, Guendisch D, Collier AC, Bingham JP. Conotoxin truncation as a post-translational modification to increase the pharmacological diversity within the milked venom of Conus magus. Toxicon. **2013** Aug 1;70:170-8.

AWARDS

Na Kama Kai Volunteer of the Year Award (**2012**)

Tribal Membership Initiative Fellowship (**2012-2014**)

UCSD Graduate Diversity & Outreach Award (**2014)**

National Geographic expedition granted Finalist (**2014**)

Waitt Foundation Rapid Ocean Conservation award (**2014**)

Creator of ‘Ike: Wisdom to Whisper $6000 Kickstarter.com documentary campaign (**2014**)

Acres of Diamonds Minority Training Research Forum Oral Presentation Award (**2015**)

180-Ocean Challenge Finalist (**2016**)

Scripps Institution of Oceanography Program of Interdisciplinary Environmental Research Fellow (**2015-16**)

UCSD Global Science Policy Fellow (**2016-present**)

UCSD International Forum on Environmental Research Group (**2015-present**)

3rd place finish for the 2016 Biomimicry Global Design International Challenge (**2017**)

AAAS Mass Media Fellow (**2018**)

ABSTRACT OF THE DISSERTATION

A multi-cultural characterization of human-environmental interaction through metabolomic and microbiome profiling

by

Clifford A. Kapono

Doctor of Philosophy in Chemistry
University of California, San Diego, 2018
Professor Pieter C. Dorrestein, Chair

Hawaiian culture has long described the value of environmental observation in regards to prolonging human survival. As the human population approaches a staggering 9 billion by the end of the century, human-environmental interaction becomes an important field of study which can provide new insights towards economic, ecological and social management. Advancements in metabolomic and microbiome profiling technology has provided a novel methodology to characterize human-environmental interactions on large scales. The work presented in this dissertation merges indigenous culture and institutionalized science research in order to provide novel methods in molecular exploration.

The first chapter traditionally introduces the author and the importance of Hawaiian culture and institutionalized modern science in today's society. Current metabolome and microbiome profiling methodologies are subsequently

introduced in the opening chapter. Chapter 2 describes the spatial mapping of molecular features within the built environment. The microbiome and metabolome of four human volunteers were characterized and mapped across an office space. Annotations made across the metabolome and microbiome provided interesting lifestyle clues to each individual and also help correlate individual to particular sites within the office. Chapter 3 investigates the interactions between three species of Hawaiian coral across a benthic community. Molecular differences across *M. capitata, M. flabellatta,* and *P. lobata* were observed. Variance among healthy and disease tissues of *M. capitata* were also described. Coral molecular features were then mapped onto 3D models, offering new potential for inclusion of 3d molecular profiling into long-term ecological surveys. Chapter 4 explores the relationship between humans and the marine environment by characterizing the metabolome and microbiome of surfers. In additional to identifying that skin is a useful biome to characterize human environmental interaction, the abundance of *Psychrobacter* on surfer skin in a dose dependent manner to surfing frequency offers new insights towards how our environment shapes our microbiome. Data from the surfer biome project suggests that increase sunscreen use decreases the microbial diversity of human skin. Metabolomic profiling was able to identify key ingredients of sunscreen such as octocrylene which can be further investigated for its effects on microbial diversity. Lastly, chapter 5 describes the future potential for the presented research as well as the importance of science communication.

1. Chapter 1: Introduction to the Dissertation

1.1 Indigenous perspective

1.1.1 A Hawaiian Introduction

Both my Father and Grandfather were born in the territory of Hawaii. My Great Grandfather was born in the Kingdom of Hawaii. His father's before him were born into chiefdoms across ka pae aina o Hawaii before arriving 90 generations prior on a fleet of seven canoes during the second migration across ka moana nui a kea. They came from a place known today as the “valley of the caterpillar” where they are said to have lived for 900 generations. I was born in the district of Nuuanu in the Queen's Medical Center founded by her majesty Queen Emma Kalanikaumakaamano Kaleleonalani Naea Rooke. From birth, I was taken from Nuuanu by my parents to the eastern shores of Oahu where I spent the first years of my life next to lake Kaelepulu in the district of Kailua. It was after my 5th birthday that my family moved to a region on Hawaii island known today as Hilo paliku. I was raised in the waters of Honolii, adjacent to our home in Kauhiula. My name is Clifford Allen Kapono and I am the first generation of my Hawaiian family to be born into US citizenship. The purpose of this dissertation is to provide a cumulative body of work that exemplifies traditional Hawaiian intelligence across contemporary scientific space.

1.1.2 The importance of observational studies to Hawaiian culture

The Hawaiian proverb, i nana, a ike is famously recognized throughout Hawaii and illustrates the importance of “through observation, one learns.” Until relatively

recently, this perspective has existed as a primary driver of Hawaiian society for thousands of years. The consequences of a rapidly evolving social system merged with the complexities of global development have resulted in unprecedented trauma to both the land and its people. Recognizing the current condition of environmental spaces and human health within Hawaii communities, this dissertation proposes innovative methods that can improve both aspects of Hawaiian life. Moreover, this dissertation looks to utilize metabolomic and microbiome profiling as a means to better understand the connection between humans and their environments.

1.2 Metabolomic profiling

1.2.1 Current themes in metabolomics

One of the greatest contributions to understanding the natural world has been the ability for investigators to interpret biochemical pathways and their associated biological function[1]. Biological pathways are often affected by the chemical transformation of small molecules during metabolism, which are commonly referred to as metabolites[2]. These metabolites provide chemical clues into the dynamic phenotypes of a variety of organisms and environments. One such example of how biochemical understandings have transformed public health services is by the discovery of essential nutrients and the ability to overcome such dietary deficiencies through supplemental additives[3]. In addition, by identifying relevant biochemical targets that are associated with key biochemical systems, significant advancements in modern pharmaceutical, pesticide and biotechnology industries has been achieved resulting in massive contributions in the health[4,5], environmental [6] and agricultural sector.

Metabolite profiling, or metabolomic analysis, has been ongoing for decades as mechanisms to identify unique biomarkers for organism health[7]. Recently, efforts have been established to utilize such metabolomic profiling techniques to identify a series of biomarkers that constitute the complex interactions within ecosystems. Methods in NMR, mass spectrometry and chromatography analytical systems have provided effective mechanisms for characterizing the chemical space[8]. Metabolomic profiling of organism environment interactions provide increasing benefits for assessing organism function and health at the molecular level[9]. In turn such profiling also provides investigators with unexpected relationships and metabolite responses, which can lead to the development of new hypothesis. Because of this, metabolomics is being accepted in a number of applications in identifying organism responses to modulations in environmental change such as temperature, climate, predation, infection and habitat loss[10].

1.3 Untargeted metabolomics

1.3.1 Current methods in untargeted metabolomics

Targeted metabolomic profiling refers to the method of study where a particular class of molecules or specific pathway of interest is investigated from an existing list[2]. Recent advancements in NMR and MS-based techniques have largely facilitated the enhanced ability to characterize chemical profiles especially in previously annotated systems and studies in drug metabolism pharmacokinetics[11–13], therapeutic research[14–16], and biomarker discovery[17,18] are examples of the importance of a targeted metabolomic research. Still, higher sensitivity, improved metabolite discrimination,

coverage of the metabolome space, and modularity to perform compound-class-specific analysis of MS-based profiling serves as a huge motivating factor when choosing MS-based profiling over many of the current NMR-based approaches[17,19]. In addition the numerous benefits of MS-guided targeted metabolomic studies, untargeted profiling also serves as a new and exciting method of biological inquiry especially when exploratory measures are required[20–22].

Advantages in untargeted approaches in metabolomics include the ability to simultaneously measure thousands of metabolites from biological samples with minimal experimental bias[2]. As previously mentioned, technologies such as nuclear magnetic resonance (NMR)[23,24], liquid chromatography mass spectrometry(LC-MS)[21,22], gas chromatography mass spectrometry (GC-MS)[25,26], raman spectroscopy[27] and infrared spectroscopy[28] provide researchers the ability to characterize the chemical space even with limited knowledge of the biochemical pathways of a system. Combinatorial methods such as LC-MS have proved to be extremely beneficial for the detection of many metabolites therefore being adopted in many of the studies listed here. LC-MS based approaches allow for reproducible, highly sensitive and relatively high throughput profiling in a broad range of biological systems[27,29–31].

1.3.2 Sampling

Sampling methods for untargeted metabolomic studies can often range from having hundred to thousands of smapling sites and mechanically differentiate based on the biological system. For instance where cotton swabbing of human skin surfaces[32] may provide enough biomass for large scale profiling, coral reefs may require upwards

of 10mg of coral[33] for any metabolites to be detected. Handling and storage practices should also be considered as metabolites may undergo decomposition or transformation during transport. Extractions methods for such samples can range through different solvent polarities and have significant consequences on the type of molecules detected. Such extraction methods may include, but not limited to, liquid-liquid extraction (LLE), solid-liquid extraction (SLE), solid phase extraction (SPE), microwave assisted extraction, accelerated solvent extraction, protein precipitation and more[34].

1.3.3 Data acquisition

1.3.3.1 Chromatography

Chromatography is commonly coupled with mass spectrometry in order to reduce molecular complexity and separate molecules within a sample. High performance liquid chromatography (HPLC) is a powerful method widely implemented throughout the scientific community. By incorporating C18 columns into chromatography, investigators can separate molecules belonging to a wide range of semi-polar molecules such as phenolic acids, flavonoids, glycosylated steroids, alkaloids and other glycosylated species. Conversely, by using hydrophilic interaction liquid chromatography (HILIC), more polar compounds such as sugars, amino sugars, amino acids, vitamins, carboxylic acids and nucleotides can be better separated. Moreover, advancements in chromatography such as the development of ultra-performance liquid chromatography have enhanced resolution of peak detection and increased throughput capabilities[35–37] .

1.3.3.2 Ionization

For detection to occur, molecules must be ionized. After separation by LC, molecules are introduced to an ion source. Electron ionization accelerates electrons towards a particular molecule resulting in a gain of charge. Alternatively, chemical ionization utilizes chemical reactions to apply charge to a particular molecule through the addition or subtraction of protons. With the advancement of atmospheric ionization methods such as atmospheric pressure ionization (API), electrospray ionization (ESI) and atmospheric chemical ionization (APCI), ions are created in the gas phase [38]. These “soft” forms of ionization reduces the amount of in-source fragmentation of “unknown” molecules and lends itself advantageous in untargeted analysis. Once ionized, molecules migrate through an analyzer.

1.3.3.3 Analyzer separation

MS analyzers separate ions based on their mass to charge ratio (m/z). Mass range, mass resolution, resolution, ion transmission and sensitivity are all parameters that are considered during the separation process. Common analyzers used in metabolomics include time of flight (TOF) analyzers, quadrupole analysers, ion traps, and orbitraps. Although the acceleration and separation mechanisms differ, hybrid variations of these analyzers can be combined in order to work in tandem. Subsequent fragmentation of isolated molecules (MS/MS) can then be performed and sent to the detector for inventory[39–41].

1.3.3.4 Detection

Mass detection is the final stage of mass spectrometry and amplification is a important step. Signals are often very small once received in the detector and some form of electron multiplier is used[42]. During data acquisition both retention time (RT) and corresponding mass to charge ratios (m/z) can used as identifiers of unique molecules. MS/MS fragmentation can also be collection during data acquisition.

1.3.4 Data Processing

Unlike targeted metabolomic profiling, untargeted approaches are highly complex and usually very computationally dense. Large data storage and technical processing is usually required even prior to analysis. Outputs can be exported on the orders of gigabytes per sample and often require large bandwidths for data transfer. With each file containing thousands of potential metabolite information, innovative analytical tools are often used for analysis. Many MS manufacturers offer data analysis tools without the need for data conversion, but these data types often cannot be compared between data acquired on different instruments. Over the last 5 years, 3rd party MS analysis plug-ins has overcome such issues and increased analytical flexibility by simplifying the ability to analyze datasets. For example, open-sourced programs such as XCMS[43,44], Metalign[45], MZmine 2, and GNPS[46] can utilize select manufacturer data formats and common post-processed file type known as mzXML, mgf, and MZml as inputs for analysis.

Advantages that Metalign offers users in the ability to convert data to and from manufacturer formats including netCDF from both LC/MS and GC/MS systems;

preprocessing (baseline correction, denoising, accurate mass calculation and export of results to manufacturer formats which help in visualization); alignment at low mass and high mass resolutions; export of univariate statistical selections to differential MS data files; export to spreadsheets for multivariate statistical analysis and conversation of multivariate statistical selection to a MS data file[47].

XCMS offers user-friendly solutions to establish visualization, statistical analysis, pathway analysis, stable isotope analysis, and metabolite identification[44]. XCMS online utilizes novel nonlinear retention time alignment, matched filtration, peak detection and peak matching. Even without internal standards, this method can identify hundreds of endogenous metabolites and use them as standards calculating nonlinear retention time correction profiles for each sample. The relative metabolite ion intensities are then directly compared to identify changes or potential biomarkers[43].

Recent improvements in the MZmine 2[48] software increases modularity and flexibility in the interface allowing for easier peak detection, alignment, identification, normalization, visualization and statistical analysis while older open-source software such as Open MS[49] continue to evolve user flexibility across multiple processing languages.

1.3.5 Data visualization

Although many of the previously open-source software previously mentioned provides users with sophisticated and robust statistical visualization methods, the field is constantly providing alternatives for innovative visualization of untargeted metabolites. Alternative visualization methods of metabolomic data can also be implemented such as

MEDI self-organizing map approach. This approach utilizes centroided and aligned retention time and mass spectral profiles which can be derived from the previously mentioned software and introduces this normalized data into self-organizing map (SOM) algorithms. In this workflow, SOM algorithm is used to assign detected features into a self-organizing grid with dimensions predefined by users. Heat-map like grids illustrate the relative abundances of detected features through a color visualization schema. Regions of interest may then be subjected to further analysis and identification via database searches of RT, high mass accuracy MS1 and MS/MS fragmentation similarities[50].

Optimus utilizes state-of-the-art feature detection algorithms from OpenMS to formulate a global depiction of metabolites from LC/MS datasets. In addition to feature detection, quantification, noise reduction, mz/rt matching and spatial mapping are also available. 3D and 2D heat-map-like spatial mapping of metabolomic signature abundances onto structural models is made possible through the Ili online software[51].

GNPS molecular networking[46] visualization platform utilizes MS/MS fragmentation patterns to identify relationships within the metabolome. Based on the fragmentation similarities, networks are created to visualize the similarities of detected metabolites and their identification. Although GNPS doesn't currently offer statistical tests such as PCOA, Random Forest, Kruskal Wallis etc, GNPS does output a spreadsheet file bucket table list that can be introduced into statistical analytical tools such as Qitta[52]. Alternatively, outputs from GNPS can also be visualized in cytoscape

where molecules are represented as nodes connected by edges based on MS2 fragmentation[53].

1.3.6 Identification of metabolites

The metabolomics standards initiative has proposed a 4 tier system of molecular annotation through mass spectrometry. Level 1 identification requires a minimum of two independent and orthogonal datasets that can be authenticated by known standards analyzed under identical experimental conditions. Level 2 identification is issued when literature or comparative data outside of the laboratory is used. Level 3 identification uses MS/MS spectral matches to putatively identify molecules and molecular families. Level 4 identification is used to apply a classification of molecules that are although unidentified or unannotated by molecular class, they are quantifiable and distinguishable from each other[54,55].

GNPS molecular networking issues both level 3 and 4 molecular annotation through spectral match comparisons from experimental spectra to known spectra of metabolites from the GNPS database. The GNPS database includes reference libraries from the Critical Assessment of Small Molecule Identification[56] (CASMI), Human Metabolome Database[57] (HMDB), MassBank[58], MassBank EU[59], MassBank of North America[60] (MONA) and National Institute of Standards (NIST). GNPS is currently the largest repository of MS/MS based metabolite.

1.4 Microbiome profiling

1.4.1 Current themes in microbiome research

Microbiome research aims to understand biological systems through the exploration of organism microbiota. These communities are comprised of hosts, bacteria, fungi, protists, and viruses. Since the initiation of the Human Microbiome Project in 2007, microbiome research has grown exponentially. Recent studies linking human allergies, brain function, diet and even behavior to gut microbiota have highlighted the value of exploring host-microbiome interactions. Through the In order to further investigate such relationships, it is critical to develop improved methodologies when collecting microbiome inventories and establish taxonomic baselines which can be used in future studies.

1.4.1.1 The Earth Microbiome Project

Though the appreciation for microbes continues to grow, few studies have looked at understanding the primary structure of microbial communities across the globe. The Earth Microbiome Project[61] (EMP) is a global effort to inventory massive amounts of microbial diversity across the planet. Within four years of the EMP's initiation in 2010, the project soon had over 200 collaborators that provided samples from over 40 biomes from the deep sea, lakes, and remote human populations. By 2017, datasets have increased by nearly 100 fold[62] and the first large scale meta analysis of the project was performed uncovering unprecedented observations towards how environments, samples sites, organism type and geography affect microbial abundance and diversity[63].

1.4.1.2 The American Gut Project

Analogous to the EMP, one such study that looks to better understand the relationship humans have with their associated microbiota[64,65] is the American Gut Project[64] (AGP). The AGP is a part of the EMP and specifically looks to understand the roles gut bacteria have on human health. Prior to the AGP, studies that investigated human gut were limited to at most hundreds of samples. The AGP has created an open platform where participants volunteer their samples thus creating a database of over 7,000 individuals since its creation in 2012. By profiling the human gut microbiome, investigators are beginning to identify correlations between active lifestyles, diet, alcohol consumption and microbial diversity.

1.4.2 Current methods in microbiome research

Both a strength and significant challenge to studies like the HMP, EMP, and AGP is the shear size and complexity of the datasets. A majority of the data is acquired across different temporal and spatial circumstances. Developing standard and robust sampling methods is critical for downstream analysis especially when mining large datasets. Standard operating procedures developed by panels of experts in the HMP, EMP and AGP have been communicated to researchers and volunteers through primary literature[61,66]. Not only do these operating procedures need to meet participant confidentiality and institutional review board review board requirements, but take into consideration government regulation of sample transport. After collection, samples and/or sampling devices are stored at -80C until DNA extraction

Despite the microbiome comprising of a consortium of microorganisms, many studies including the HMP, EMP and AGP focus primarily on bacteria DNA. After extraction, DNA is sent for 16S rRNA amplicon sequencing. Illumina HiSeq technology is often implemented and provides ultra-high throughputs of complex microbial communities[67]. Error-corrected pyrosequences can then be introduced into comparative analysis softwares such as the open source quantitative insights into microbiology' (QIIME[68]). QIIME represents a robust platform that is able to provide investigators with the ability to rapidly analyse heterogenous microbial datasets. Within the QIIME workspace, users can implement network analysis, obtained histograms of microbial diversity and determine keystone drivers of population differences.

1.5 In this Dissertation

1.5.1 Chapter 2: 3D snapshot of office

This published chapter (Kapono et. al, 2018)[69] is the first report to integrate 3D cartography of MS and 16S multi-omic profiling for forensic use. The purpose of this study was to characterize the interaction between humans and the built environment while exploring the potential for forensic application. In this study, an office space and four of its inhabitants were profiled using MS and 16S rRNA amplicon sequencing. Hundreds of samples were taken throughout the built environment as well as on the bodies of volunteers. A 3d model of the built environment was created by using an 3D systems iSense Ipad scanner[70]. Molecular and microbial features from individuals provided clues into lifestyle and medical history of volunteers. Acquired features across volunteers and the built environment were used to determine individual uniqueness as

well as how participants interacted with the built environment. From the study, “molecular fingerprints” were observed and characterize suggesting new potentials for advancing forensic technologies.

1.5.2 Chapter 3: 3D snapshot of Coral reef

This chapter applies multi-omic molecular profiling and spatial mapping technology towards the marine environment. The purpose of this study was to develop novel methodologies that characterize coral reef assemblages. Although microbiome, metabolome and 3D spatial mapping studies exist within coral research, no study has merged these technologies into a single study. 3D molecular cartography was implemented to investigate a coral assemblage in Kapoho, Hawaii in aims to better understanding recent outbreaks of coral tumor. Multivariate statistical analysis identified key molecular features that distinguish coral species as well as provide insight into the molecular world that defines individual populations. Additionally, molecular features that describe diseased coral tissue were identified. This research provides new methods to investigate coral reef assemblages as well as introduces their value for inclusion into coral reef long term ecological studies.

1.5.3 Chapter 4: Surfer Biome Project

Chapter 4 describes the science behind human environmental interaction from an indigenous perspective. The previously described methods of multi-omic molecular profiling were implemented to characterize the relationship between humans and the marine environment. The metabolome and microbiome of an international surfing cohort was profiled in efforts to determine 1) the ocean leaves a molecular fingerprints on

surfers, 2) geography of surfing populations differentiate biomes, and 3) surfers and molecularly different than non surfers. Surfers were profiled from Ireland, Morocco, London, Hawaii, San Francisco, and locally in San Diego. This project was able to characterize differences between surfers and non surfers, characterize geographic differences between different surfing populations and also identify molecular features that affect microbial diversity across surfers. Lastly, microbial features were identified that describe the frequency in which surfers interact with the ocean. This study has gained international recognition as was highlighted in The New York Times, Surfer Magazine, Hawaii Business Magazine, CBS, NBC and more.

1.5.4 Chapter 5: Future Directions

The final chapter is a brief perspective that describes how future investigations can benefit from the research of this dissertation. The technology described has great potential and application in the Hawaiian community. The value of molecular profiling is discussed and proposed to be included in the fight against fungal disease within the Hawaiian forest community. The same technology is also proposed to be used as coral researchers begin long term ecological studies in the Northwestern Hawaiian Islands. An example of how the research in this dissertation can offer new hypothesis is also described. Studies to investigate the bioactivity of surfer associated marine bacteria can be tested in order to begin understanding the potential therapeutic benefits of visiting to the ocean. Lastly, future career directions are discussed in the concluding remarks.

1.6 Acknowledgments

The UC San Diego Graduate Division provided the necessary funds needed to begin training in the Dorrestein lab. From the Tribal Initiative Fellowship, training of liquid chromatography mass spectrometry began under the mentorship of Mike Meehan, Laura Sanchez and Vanessa Phelan.

1.7 References

1. German JB, Bruce German J, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. Metabolomics. 2005;1(1):3–9.

2. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012 Mar 22;13(4):263–9.

3. Carpenter KJ, Roberts S, Sternberg S. Nutrition and immune function: a 1992 report. Lancet. 2003 Jun 28;361(9376):2247; author reply 2247–8.

4. Lenz EM, Williams R, Wilson ID. Metabonomics in the Pharmaceutical Industry. In: Metabolome Analyses: Strategies for Systems Biology. p. 337–48.

5. Abate-Shen C, Shen MM. The prostate-cancer metabolome. Nature. 2009;457(7231):799–800.

6. Bonvallot N, Canlet C, Tremblay-Franco M, Blas-Y-Estrada F, Gautier R, Cordier S. A low-dose of a complex pesticide mixture disrupts the metabolome of pregnant rats and their offspring. Toxicol Lett. 2014;229:S234.

7. Tomita M, Nishioka T. Metabolomics: The Frontier of Systems Biology. Springer Science & Business Media; 2006. 256 p.

8. Viant MR, Bearden DW, Bundy JG, Burton IW, Collette TW, Ekman DR. International NMR-based environmental metabolomics intercomparison exercise. Environ Sci Technol. 2009 Jan 1;43(1):219–25.

9. Griffiths WJ. Metabolomics, Metabonomics and Metabolite Profiling. Royal Society of Chemistry; 2008. 323 p.

10. Viant MR, Sommer U. Mass spectrometry based environmental metabolomics: a primer and review. Metabolomics. 2012;9(S1):144–58.

11. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J. Procedure

for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics. 2011;8(1):133–42.

12. Simithy J, Sidoli S, Garcia BA. Integrating Proteomics and Targeted Metabolomics to Understand Global Changes in Histone Modifications. Proteomics. 2018 Mar 7;e1700309.

13. Kumari S, Sharma P. Targeted Metabolomics Reveals Hippurate as a Urinary Potential Marker for Diabetic Nephropathy. Current Metabolomics. 2016;4(2):121–8.

14. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted Metabolomics. In: Current Protocols in Molecular Biology. 2012.

15. Hopfgartner G, Varesio E. Tandem Mass Spectrometry Hyphenated with HPLC and UHPLC for Targeted Metabolomics. In: Metabolomics in Practice. 2013. p. 21–37.

16. Cassera MB. Targeted metabolomics applied to the study of intracellular eukaryotic pathogen metabolism [Internet]. Identification and Data Processing Methods in Metabolomics. 2015. p. 126–39. Available from: http://dx.doi.org/10.4155/fseb2013.14.247

17. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner H-P. Targeted Metabolomics for Biomarker Discovery. Angew Chem Int Ed. 2010;49(32):5426–45.

18. Breit M, Baumgartner C, Netzer M, Weinberger KM. Clinical Bioinformatics for Biomarker Discovery in Targeted Metabolomics. In: Translational Bioinformatics. 2016. p. 213–40.

19. Hoffmann T, Krug D, Hüttel S, Müller R. Improving Natural Products Identification through Targeted LC-MS/MS in an Untargeted Secondary Metabolomics Workflow. Anal Chem. 2014;86(21):10780–8.

20. Vinayavekhin N, Saghatelian A. Untargeted Metabolomics. In: Current Protocols in Molecular Biology. 2010.

21. Hill CB, Roessner U. Advances in high-throughput untargeted LC–MS analysis for plant metabolomics [Internet]. Advanced LC-MS Applications in Metabolomics. 2015. p. 58–71. Available from: http://dx.doi.org/10.4155/fseb2013.14.54

22. Wehrens R, Hageman JA, van Eeuwijk F, Kooke R, Flood PJ, Wijnker E. Improved batch correction in untargeted MS-based metabolomics. Metabolomics [Internet]. 2016;12(5). Available from: http://dx.doi.org/10.1007/s11306-016-1015-8

23. Čuperlović-Culf M. Metabolomics NMR data preprocessing – analysis of individual spectrum. In: NMR Metabolomics in Cancer Research. 2013. p. 215–59.

24. Boiteau RM, Hoyt DW, Nicora CD, Kinmonth-Schultz HA, Ward JK, Bingol K. Structure Elucidation of Unknown Metabolites in Metabolomics by Combined NMR

and MS/MS Prediction. Metabolites [Internet]. 2018 Jan 17;8(1). Available from: http://dx.doi.org/10.3390/metabo8010008

25. Murgia A, Caboni P, Cadoni E, Serra M, Marongiu F, Laconi E. A GC-MS untargeted metabolomics analysis in the plasma and liver of rats lacking dipeptidyl-peptidase type IV enzyme activity. J Physiol Biochem. 2017 Nov;73(4):575–82.

26. Papadimitropoulos M-EP, Vasilopoulou CG, Maga-Nteve C, Klapa MI. Untargeted GC-MS Metabolomics. Methods Mol Biol. 2018;1738:133–47.

27. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol. 2004;22(5):245–52.

28. Ellis G, Claybourn M, Richards SE. The application of fourier transform raman spectroscopy to the study of paint systems. Spectrochim Acta A. 1990;46(2):227–41.

29. Monteiro MS, Carvalho M, Bastos ML, Guedes de Pinho P. Metabolomics analysis for biomarker discovery: advances and challenges. Curr Med Chem. 2013;20(2):257–71.

30. Daskalchuk T, Ahiahonu P, Heath D, Yamazaki Y. The Use of Non-targeted Metabolomics in Plant Science. In: Biotechnology in Agriculture and Forestry. p. 311–25.

31. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. Metabolites. 2012 Oct 18;2(4):775–95.

32. Bouslimani A, Porto C, Rath CM, Wang M, Guo Y, Gonzalez A. Molecular cartography of the human skin surface in 3D. Proc Natl Acad Sci U S A. 2015 Apr 28;112(17):E2120–9.

33. Garg N, Kapono CA, Lim YW, Koyama N, Vermeij MJA, Conrad D. Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures. Int J Mass Spectrom. 2015;377:719–27.

34. Zhou B, Xiao JF, Tuli L, Ressom HW. LC-MS-based metabolomics. Mol Biosyst. 2012 Feb;8(2):470–81.

35. Lu W, Bennett BD, Rabinowitz JD. Analytical strategies for LC-MS-based targeted metabolomics. J Chromatogr B Analyt Technol Biomed Life Sci. 2008 Aug 15;871(2):236–42.

36. MacNair JE, Lewis KC, Jorgenson JW. Ultrahigh-Pressure Reversed-Phase Liquid

Chromatography in Packed Capillary Columns. Anal Chem. 1997;69(6):983–9.

37. Guillarme D, Nguyen DT-T, Rudaz S, Veuthey J-L. Recent developments in liquid chromatography—Impact on qualitative and quantitative performance. J Chromatogr A. 2007;1149(1):20–9.

38. Dams R, Huestis MA, Lambert WE, Murphy CM. Matrix effect in bio-analysis of illicit drugs with LC-MS/MS: influence of ionization type, sample preparation, and biofluid. J Am Soc Mass Spectrom. 2003 Nov;14(11):1290–4.

39. Kebarle P, Tang L. From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. Anal Chem. 1993;65(22):972A – 986A.

40. Michalski A, Cox J, Mann M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. J Proteome Res. 2011;10(4):1785–93.

41. de Hoffmann E. Mass Spectrometry. In: Kirk-Othmer Encyclopedia of Chemical Technology. 2005.

42. Daly NR. Scintillation Type Mass Spectrometer Ion Detector. Rev Sci Instrum. 1960;31(3):264–7.

43. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. Anal Chem. 2012;84(11):5035–9.

44. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006 Feb 1;78(3):779–87.

45. Lommen A. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem. 2009 Apr 15;81(8):3079–86.

46. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016 Aug 9;34(8):828–37.

47. Lommen A, Kools HJ. MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. Metabolomics. 2011;8(4):719–26.

48. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11(1):395.

49. Röst HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat

Methods. 2016 Aug 30;13(9):741–8.

50. Goodwin CR, Covington BC, Derewacz DK, McNees CR, Wikswo JP, McLean JA. Structuring Microbial Metabolic Responses to Multiplexed Stimuli via Self-Organizing Metabolomics Maps. Chem Biol. 2015 May 21;22(5):661–70.

51. Protsyuk I, Melnik AV, Nothias L-F, Rappez L, Phapale P, Aksenov AA. 3D molecular cartography using LC-MS facilitated by Optimus and ’ili software. Nat Protoc. 2018 Jan;13(1):134–54.

52. Qiita [Internet]. [cited 2018 May 1]. Available from: https://qiita.ucsd.edu/

53. Sanchez LM, Phelan VV, Wang M, Bandeira N, Dorrestein PC. Visualizing diverse chemical families with molecular networking. Planta Med [Internet]. 2015;81(11). Available from: http://dx.doi.org/10.1055/s-0035-1556181

54. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA. Proposed minimum reporting standards for chemical analysis. Metabolomics. 2007;3(3):211–21.

55. Creek DJ, Dunn WB, Fiehn O, Griffin JL, Hall RD, Lei Z. Metabolite identification: are you sure? And how do your peers gauge your confidence? Metabolomics. 2014;10(3):350–3.

56. Oberacher H. Applying Tandem Mass Spectral Libraries for Solving the Critical Assessment of Small Molecule Identification (CASMI) LC/MS Challenge 2012. Metabolites. 2013 Apr 26;3(2):312–24.

57. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N. HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007;35(Database):D521–6.

58. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010;45(7):703–14.

59. Slobodnik J, Dulio V. NORMAN Association: A Network Approach to Scientific Collaboration on Emerging Contaminants and their Transformation Products in Europe. In: Transformation Products of Emerging Contaminants in the Environment. 2014. p. 903–16.

60. [No title] [Internet]. [cited 2018 May 1]. Available from: http://mona.fiehnlab.ucdavis.edu/

61. Gilbert J. The Earth Microbiome Project: A new paradigm in geospatial and temporal studies of microbial ecology [Internet]. SciVee. 2012. Available from: http://dx.doi.org/10.4016/46411.01

62. Gilbert JA, Jansson JK, Knight R. Earth Microbiome Project and Global Systems

Biology. mSystems. 2018;3(3):e00217–17.

63. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017 Nov 23;551(7681):457–63.

64. Debelius JW, Xu Z, Vázquez-Baeza Y, Knight R, Wolfe E, McDonald D. Turning Participatory Microbiome Research into Usable Data: Lessons from the American Gut Project. J Microbiol Biol Educ. 2016;17(1):46–50.

65. Knight R. Follow Your Gut: The Enormous Impact of Tiny Microbes. Simon and Schuster; 2015. 128 p.

66. Forum F, Food and Nutrition Board, Institute of Medicine. The Human Microbiome, Diet, and Health:: Workshop Summary. National Academies Press; 2013. 196 p.

67. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012 Aug;6(8):1621–4.

68. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010 May;7(5):335–6.

69. Kapono CA, Morton JT, Bouslimani A, Melnik AV, Orlinsky K, Knaan TL. Creating a 3D microbial and chemical snapshot of a human habitat. Sci Rep. 2018 Feb 27;8(1):3669.

70. Ukida H, Tanimoto Y, Sano T, Yamamoto H. 3D object shape and reflectance property reconstruction using image scanner. In: 2009 IEEE International Workshop on Imaging Systems and Techniques [Internet]. 2009. Available from: http://dx.doi.org/10.1109/ist.2009.5071610

2. Chapter 2: Office Study

2.1 Introduction to Chapter 2

With a human global population reaching upwards of 9 billion by 2050[1], anthropogenic development continues to increase. Additionally, the human built-environment represents a new ecological environment and one filled with a diversity of human-derived molecular features. Defining the chemical and microbial associations of these human-built environments, where we spend most of our time, provide researchers with new interpretations of how we as a species may be influencing the overall ecology or our planet.

Molecules, whether from microbes or from environmental factors, present investigators with new opportunity to understand microbial habitats within the built environment. Microbial inventories of habitats including hospitals[2–5], households[3,6–10], space[11–13], and bathrooms[14–17] demo[18]nstrate the diversity of indoor microbial communities, and how their occupants influence the microbial composition[19,20]. Among studies that characterize harmful indoor chemicals such as phthalate[21–23], bisphenyl A[24,25], and metals[26–29], we have little insight into the diversity of molecules present in the human habitat, or how such molecules relate to the human behavior.

One of the earliest examples of molecular cartography was in the mapping of antigenic sites on the surface of proteins[30]. Since then, several studies have used multi omic techniques to map molecular features on the surfaces of various subjects[31–34]. 3d molecular cartography serves as a novel tool to interpret spatial distributions of key molecular features within systems. Chapter 2 looks to explore the human sources of

molecular features or “fingerprints” throughout the built environment to better understand the fundamentals of human-environmental interaction. By developing observational methods to profile the built space and visualize these findings in 3d, this chapter also proposes new advancements in forensic science.

In Hawaii, understanding the origins of resources and people are believed to be very important[35]. This perspective is exemplified by the traditional saying, *nana i ke kumu* or “look to the source.” This project offered a contemporary application to my ancestral teachings as was a perfect study to develop the skills necessary for studies described in chapter 3 and 4. One particular result obtained from the molecular profiling study was identifying the sources of molecules based on volunteer biology and chemistry. This study reinforced the contemporary value traditional Hawaiian perspectives have in science research.

2.2 Chapter 2 appendix of: “Creating a 3D microbial and chemical snapshot of human habitat”

SCIENTIFIC REPORTS

OPEN

Creating a 3D microbial and chemical snapshot of a human habitat

Received: 11 October 2017
Accepted: 5 February 2018
Published online: 27 February 2018

Clifford A. Kapono[1], James T. Morton[3,7], Amina Bouslimani[2], Alexey V. Melnik[2], Kayla Orlinsky[3], Tal Luzzatto Knaan[2], Neha Garg[2], Yoshiki Vázquez-Baeza[3], Ivan Protsyuk[4], Stefan Janssen[7], Qiyun Zhu[7], Theodore Alexandrov[2,4], Larry Smarr[3,5,6], Rob Knight[3,6,7] & Pieter C. Dorrestein[2,6]

One of the goals of forensic science is to identify individuals and their lifestyle by analyzing the trace signatures left behind in built environments. Here, microbiome and metabolomic methods were used to see how its occupants used an office and to also gain insights into the lifestyle characteristics such as diet, medications, and personal care products of the occupants. 3D molecular cartography, a molecular visualization technology, was used in combination with mass spectrometry and microbial inventories to highlight human-environmental interactions. Molecular signatures were correlated with the individuals as well as their interactions with this indoor environment. There are person-specific chemical and microbial signatures associated with this environment that directly relate who had touched objects such as computers, computer mice, cell phones, desk phone, table or desks. By combining molecular and microbial investigation forensic strategies, this study offers novel insights to investigators who value the reconstructing of human lifestyle and characterization of human environmental interaction.

Humans constantly interact with their environment and unintentionally leave behind a plethora of chemical and microbial traces[1,2]. Many of these signatures are substantially dictated by their habits and lifestyle choices of the individual[3–5]. It was hypothesized that by studying molecular and microbial clues across an office space, associations between objects of that environment and its inhabitants can be made. Since it may even be possible to identify personal habits for each individual from these signatures[4], such trace evidence could tremendously benefit forensic science[6]. More importantly, it is necessary to properly document and understand how to best utilize such information. Solving this task has the potential to modernize forensic studies especially when correlating a particular individual to a surface or object within an environment is paramount to 'solving a case'. Additionally, beyond forensics applications, this investigation also addresses the growing interest in understanding the role of humans as unique reservoirs of chemicals and how their microbes contribute to the overall molecular diversity of a built environment[6].

To analyze the molecular and microbial traces of humans in a built environment, we employed surface sampling and visualized these findings with 3D molecular cartography[1,4]. In 3D molecular cartography, samples are collected across a space and the data is then mapped using their coordinates of the 3D model. The 3D model of the office with volunteers sitting at the conference desk was created using a 3D scanner (Supplemental Fig. 1). A detailed analysis of uniquely distributed molecules and microbes provided interesting clues into the lifestyles of each volunteer as previously demonstrated to individuals and their phones[4]. Using Global Natural Products Molecular Social networking[7] (GNPS), molecular networking was able to correlate office molecules and office object molecules with volunteer's person or personal care products. 16S rRNA amplicon sequencing was used to inventory bacteria taxonomy across volunteers and the office. 16S rRNA amplicon sequencing provides insights

[1]Department of Chemistry, University of California San Diego, La Jolla, CA, USA. [2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA, USA. [3]Department of Computer of Science and Engineering, University of California San Diego, La Jolla, CA, USA. [4]Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117, Heidelberg, Germany. [5]California Institute for Telecommunications and Information Technology, University of California San Diego, La Jolla, CA, USA. [6]Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA. [7]Department of Pediatrics, University of California San Diego, La Jolla, CA, USA. Correspondence and requests for materials should be addressed to R.K. (email: rknight@ucsd.edu) or P.C.D. (email: pdorrestein@ucsd.edu)

into the bacteria that are present in a sample. Liquid chromatography mass spectrometry is an analytical technique that separates and detects molecules based on their hydrophobicity and atomic weight. LC-MS/MS was used to inventory the molecules in the room. Such associations provide insights into the lifestyles of volunteers in the built environment. We further demonstrated that partial least squares analysis[8] (PLS) can be used to distinguish volunteers' skin metabolome and skin microbiome with each other and also with the biomes of the built environment. Lastly, we demonstrate that SourceTracker V.1[9] can be used to identify the relationships between individuals and the overall molecular and microbial profiles observed in the environment.

Results

The office environment included a desk, bookshelf, conference table, four chairs, garbage and recycling cans, the carpet-covered floor, phones, computers and computer accessories and walls. Four individuals inhabited the office environment for a one-time sample collection and were labeled as volunteers 1–4 (Supplemental Fig. 1). Volunteer 3 is the main inhabitant of the office; volunteer 1 had only been in the office twice in the past two years while volunteers 2 and 4 frequented the office 2–4 times a month for short periods (minutes to hours) of time during each visit. A total of 390 sites were sampled from the office environment; 276 sites were sampled on the skin, clothes, and accessories of volunteers. Each volunteer was sampled 33 times throughout his or her head, face, arms, legs, chest, and hand. Each built environment site was sampled between 2–17 times. To obtain both chemical and microbial inventories, each sample site was sampled twice within 1 cm apart making sure not to overlap extraction points. One swab from each sample site was subjected to untargeted liquid chromatography tandem mass spectrometry (LC-MS/MS) for detection of small molecules and the other swab was subjected to 16S rRNA amplicon sequencing for assessing the microbial taxa. The resulting data was overlaid onto the 3D model in the 'ili web application[10]. No particular order was assigned during sample collection. The volunteers also provided 27 personal care products such as shampoo, toothpaste, shaving cream, sinus medication, deodorant, body wash, aftershave and essential oils for metabolomic analysis which later aided in assigning ownership of molecules across the molecular space.

After LC-MS/MS data acquisition was performed, each sample site yielded thousands of raw spectral data points that when combined in GNPS and visualized in Cytoscape, produced a molecular network that described the chemistry of the office (Fig. 1a). A molecular network provides a visual representation of the molecules that were fragmented by tandem mass spectrometry[11]. After all sample sites were subjected to LC-MS/MS metabolomic profiling, a total of 301,230 raw spectral files were obtained. In an attempt to reduce low quality spectra, specific GNPS data analysis workflow parameters were used (min matched peak = 6, where the peaks are the ions in the spectrum after tandem mass spectrometry was performed, min cluster size = 2). During the construction of the molecular network, 21,732 spectra of the obtained 301,230 spectra were used to generate a molecular network of the office (Fig. 1a). Each node represents a unique molecule, while the edges connecting each node suggests molecular similarity. Self-looping nodes (single nodes and edges connecting to themselves), illustrate molecular uniqueness under the GNPS processing parameters. Within the network, 799 nodes (~4%), were matched to the reference spectra that make up the GNPS library. Of these nodes, 15,274 (70%) nodes were unique to the built environment (Fig. 1a). Volunteer 1, 2, 3, and 4 produced 5,903 (27.2%), 5,859 (26.9%), 5,390 (24.8%), and 5874 (27%) independent nodes respectively. The built environment shared 4,325 (19.9%), 4,154 (19.1%), 3,841 (17.6%), and 4,018 (18.4%) nodes with volunteers 1, 2, 3 and 4 respectively. 5,510 nodes were associated with personal product standards alone. Across the entire molecular network, 666 (3.1%) nodes were shared among all four volunteers, the built environment and volunteer personal care products. Figure 1g reveals representative molecules that were detected and annotated. It was also shown that each molecule was found to help with forensic analysis. Sodium lauryl sulfate (Fig. 1b), avobenzone (Fig. 1f), and D-erythro-sphingosine are personal care molecules shared between the inhabitants and the built environment. Theophylline (Supplemental Fig. 2c), DEET (Fig. 1e), and azoxystrobin (Fig. 1h) were molecules shared between the inhabitants and the environment. Amlodipine (Fig. 1d) was only found on the volunteers, while nobiletin (Fig. 1g) was found exclusively in the built environment

Mapping the location of microbes and molecules. Molecular and microbial annotations could assign clues about the lifestyles of each inhabitant of the office space. The majority of annotated molecules in the office space were traced back to the volunteer's individual habits and daily routines. These results are consistent with our previous findings that the majority of molecules from lifestyles (i.e. personal care products, medications, food-derived molecules) in the chemical composition are derived from the human skin surface[12] and that they can also be found on objects we touch[4]. Global Natural Products Social molecular networking[7] provided level two annotations for some of the detected ions in this study, in accordance with the 2007 metabolomics standard initiative[13]. Relative mass spectrometry feature intensities and sub-Operational Taxonomic Unit (sOTU) abundances were further plotted in a recently introduced 3D cartography tool called 'ili[10] (Fig. 2a–p). Avobenzone (Fig. 2a), a major active ingredient in sunscreen, was found on many of the volunteers' faces, hands, table, phones and computers, but was most intense on the computer, face and chest of volunteer 4. Diethyltoluamide (DEET) (Fig. 2b) was most abundant on the phone of volunteer 3 and also around the desk he was sitting next to. Laureth sulfate (Fig. 2c), a shampoo ingredient, was found on all volunteers as well as on the table. Amlodipine (Fig. 2d), also known as Norvasc™, is a calcium channel blocker used to treat high blood pressure, and was found primarily on the face and hands of volunteer 1 and the objects sitting in front of this person. A D-erythro-sphingosine (17:1) (Fig. 2e), a lipid produced by eukaryotes, was found in high abundance on the faces of all volunteers, and the flavonoid nobiletin (Fig. 2f) found in citrus was found in highest abundance on the computer and face of volunteer 4. The same molecule was also found in the lip balm of personal care product data set. Theophylline (Fig. 2g), a metabolism by product of caffeine, was found to be present on three of the volunteers although most intense on the face and hands of volunteers 1 and 2. Finally, azoxystrobin (Fig. 2h), is a known antifungal and

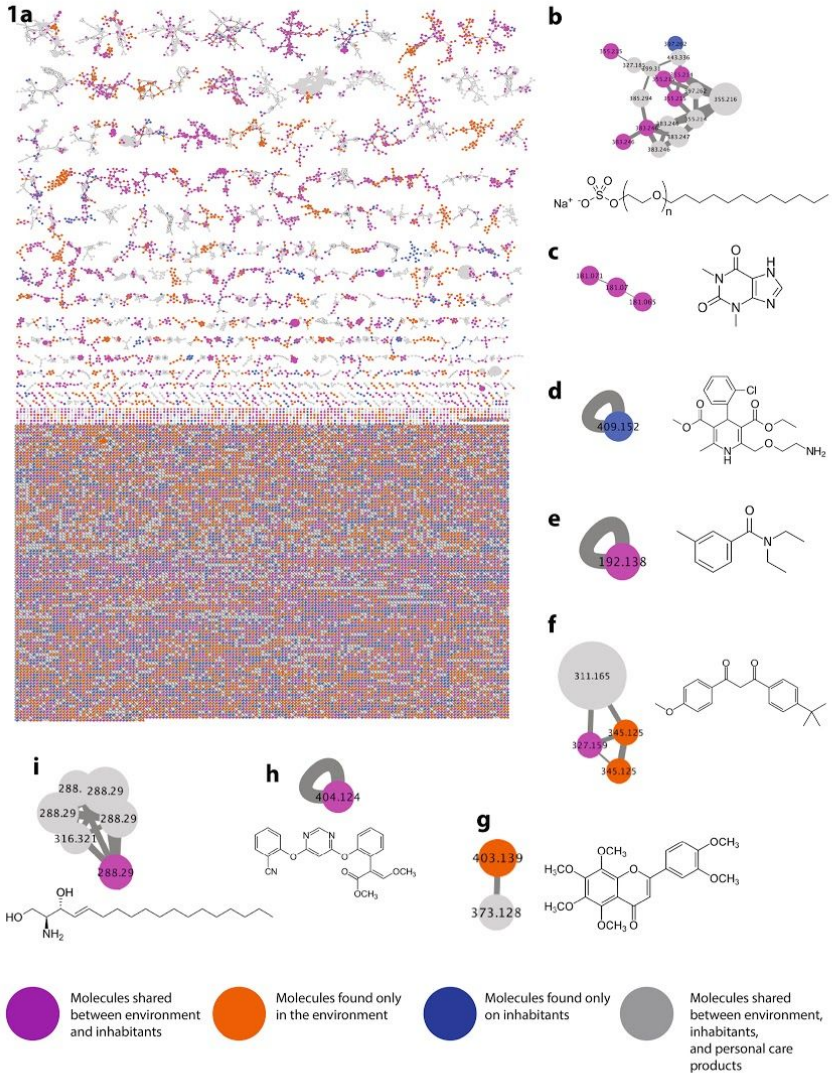


Figure 1. A molecular network of the office including its inhabitants and personal care products. The colors of the nodes illustrate that molecules are found on either the volunteers, built environment, or from personal care products of volunteers. (**a**) Blue nodes represent molecules exclusively found on volunteers. Orange nodes represent molecules exclusively found on environment. Pink nodes represent molecules shared between people and environment. Grey nodes represent molecules between environment, inhabitants and personal care products. (**b**) Sodium Lauryl Sulfate, *m/z* 397.262. (**c**) Theophylline *m/z* 181.071. (**d**) Amlodipine medication, *m/z* 409.152 (**e**) DEET, *m/z* 192.138 (**f**) Avobenzone, *m/z* 311.165 (**g**) Nobiletin, *m/z* 403.139 (**h**) Azoxystrobin, *m/z* 404.124 (**i**) D-erythro-sphingosine, *m/z* 288.29.

previously described as a common food contaminant[14] that is detected exclusively on the hand, computer and phone of volunteer 2.

Similarly, the spatial mapping of V4 16S rRNA amplicon data[15,16] in the office space revealed many distinct localizations of specific bacterial communities. In total 44,393 unique sOTU were observed in this study. The soil bacteria *Nocardiaceae* (Fig. 2i), was abundant on the floor of the office. The environmental bacterium, *Acinetobacter guillouiae*, is often found in water, soil, and also feces[17] (Fig. 2j). This bacterium was observed on

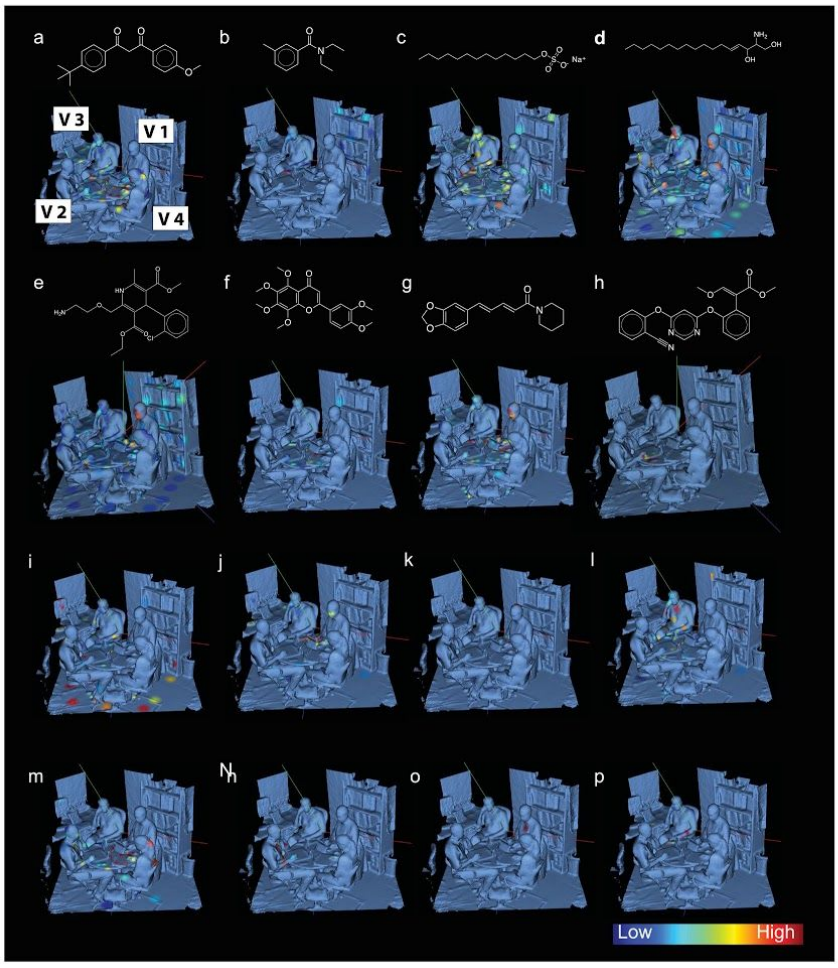


Figure 2. Distribution of molecules and microbial taxa in a human habitat and its volunteers 1–4. The MS data of the office environment sampling locations and relationships to the microbes in those same locations. (**a**) Avobenzone, (**b**) DEET, (**c**) Sodium Laureth Sulfate, (**d**) C17 Sphingosine, (**e**) Amlodipine (Norvasc), (**f**) Nobiletin, (**g**) Theophylline, (**h**) Azoxystrobin, (**i**) *Nocardiaceae*, (**j**) *Acinetobacter guillouiae*, (**k**) *Rhizobiales*, (**l**) *Synechococcus*, (**m**) *Actinomycetales*, (**n**) *Staphylococcus*, (**o**) *Veillonella parvula*, (**p**) *Chitinophagaceae*.

the hands and face of volunteer 2. An unannotated species of Gram-negative nitrogen-fixing plant-associated[18] α-proteobacteria from the order *Rhizobiales* (Fig. 2k) was found on volunteer 1's head, hands and upper back. A marine Cyanobacteria from the genus *Synechococcus* (Fig. 2l) was found on the body of volunteer 3. Actinomycetales (Fig. 2m) was found in high abundances on the body, phone and computer of volunteer 4. A bacteria from the genus *Staphylococcus* (Fig. 2n) was found to be specific to volunteer 2. A microbe associated with the human oral microbiome, *Veillonella parvula* (Fig. 2o), was found on the arms and chest of volunteer 1. Finally, *Chitinophagaceae* (Fig. 2p), a soil associated bacterium, was localized to the hands and phone of volunteer 3.

Distinguishing environmental inhabitant "biome" uniqueness. Once unique distributions of the molecules and microbes were established, a partial least squares[8] (PLS) biplot was constructed to assess if the molecular "biomes" of each inhabitants were statistically distinguishable (Fig. 3). Partial least squares (PLS) was used for data representation as it was reported to perform well in the case of a low number of samples and a high number of dimensions (molecular features and microbial taxa)[8]. Molecular features are identified by inputting LC-MS/MS spectral data into the OpenMS[19]. Molecular features are defined as molecule's atomic weight through

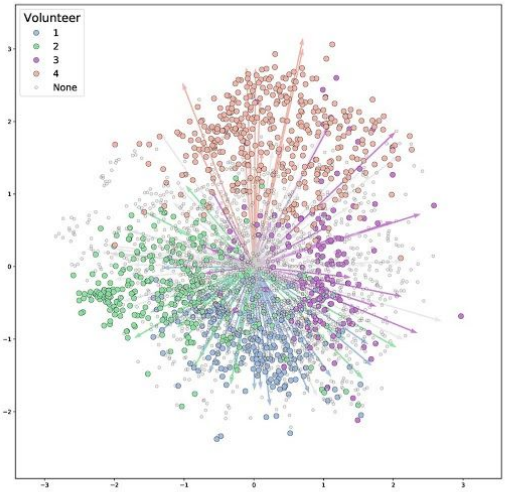


Figure 3. Partial least squares statistical biplot of volunteers and the built office space environment. Arrows represents unique microbes while circles represent unique molecules. Colors assign what unique volunteer the microbe or molecule was isolated from. Volunteer 1 is represented by the color blue, volunteer 2 is represented by color green, volunteer 3 is represented by color pink, and volunteer 4 is represented by color salmon. Color Grey illustrates microbes or arrows that cannot be correlated to a volunteer.

mass to charge ratio (*m/z)* and retention time (RT). RT reflects the hydrophobicity property of the molecule. By combining these defining properties, molecular feature detection provides added confidence in distinguishing unique molecules from a complex mixture. Overlapping arrows (representative of microbes) and circles (representative of molecules) of a particular color illustrate that there is a correlation between the microbes and molecules within a volunteer. Volunteer 1 is represented by the color blue, volunteer 2 is represented by color green, volunteer 3 is represented by color pink, and volunteer 4 is represented by color salmon. Circles and arrows (microbes and molecules) that could not be matched to a volunteer are represented in gray. The overlapping of arrows and circles (microbes and molecules) of a single color demonstrate that there are unique correlations between the microbes and the metabolome of all volunteers. Furthermore, the top three molecular features in the PLS biplot were detected through OpenMS[19] and microbial taxa through SortMeRNA[20]. Using this technique we were able to uniquely identify 41, 21, 23, 11 microbial reads that identify volunteers 1, 2, 3 and 4 respectively with a *p*-value less than 0.001 within this office environment. Additionally, we were able to characterize that 218, 255, 153, and 388 molecular features that uniquely identify each volunteer 1, 2, 3 and 4 respectively with *p*-value less than 0.001.

The top three microbes that had the highest PLS indicator value of 0.90, 0.83 and 0.79 for Volunteer 1 where a plant associated *Agrobacterium* bacteria, *Rhizobiales*, and *Rothia dentocariosa* oral cavity bacteria respectively (see Supplementary Table S1). The top three molecules that had the highest volunteer 1 PLS indicator value of 0.86, 0.81, and 0.78 were molecular features *m/z* 388.393 with a retention time (RT) range from 444 to 584 seconds (s), *m/z* 282.279, RT 448–583 s, and *m/z* 192.077, RT 193–203 s respectively (Supplementary Table 2). The top three microbes that had the highest volunteer 2 PLS indicator value of 0.91, 0.87 and 0.85 were *Staphylococcus* human skin bacteria, and two species of *Corynebacterium* (Supplementary Table 1). The top three molecules that had the highest volunteer 2 PLS indicator value of 0.84, 0.81, and 0.76 were molecular features *m/z* 374.326, RT 336–361, *m/z* 482.315, RT 388–413 s, and *m/z* 444.403, RT 402–445 s respectively (Supplementary Table 2). The top three microbes that had the highest volunteer 3 PLS indicator value of 0.94, 0.78, and 0.75 were two *Corynebacterium* and a match to whom the closest sequence was a *Deinococcus* extremophile respectively (Supplementary Table 1). The top three molecules that had the highest volunteer 3 PLS indicator value of 0.89, 0.87,and 0.86 were molecular features *m/z* 664.509, RT 577–601 s, *m/z* 669.466, RT 574–599 s, and *m/z* 655.995 RT 566–601 s respectively (Supplementary Table 2). The top three microbes that had the highest volunteer 4 indicator value of 0.82, 0.79, and 0.77 were the sequences that were most similar to *Deinococcus* bacteria, *Xanthomonadacae*, and *Corynebacterium* respectively (Supplementary Table 1). The top three molecules that had the highest volunteer 4 PLS indicator value of 0.96, 0.90,and 0.84 were unannotated molecular features *m/z* 368.424, RT 465–603 s, *m/z* 340.393, RT 444–599 s, and *m/z* 312.362, RT 410–575 s respectively (Supplementary Table 2). Additionally, seven microbes with statistically significant correlations to volunteer 3 (p-value < 0.001) are commonly found in marine environments (see Supplementary Table S3). These include marine species from the genera *Psychromonas, Marinomonas, Alteromonas, Vibrionaceae*, and, *Pseudoalteromonas* (Supplementary Table 1). Despite molecular annotation was limited across molecules with high correlations to specific volunteers, molecular feature *m/z* 311.164, RT 403–479 s identified as Avobenzone had a high correlation value to volunteer 4 at 0.80.

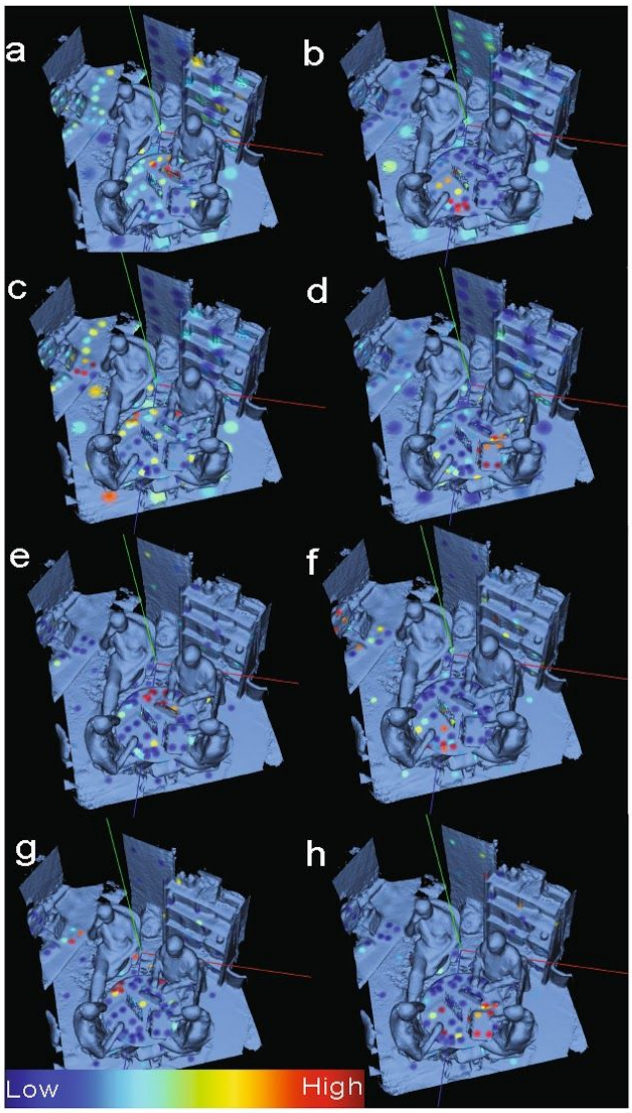


Figure 4. Distinguishing office interactions through chemical and microbial signatures using 3D cartography. Source tracking based on LC-MS profiling (**a**–**d**) as well as 16S amplicon microbial analysis (1e–h) illustrates the points and intensity of source tracking between volunteers 1–4 and samples sites across the office. Higher probability that molecular features from a sample site belong to a designated volunteer is indicated in red. Blue spots represent a lower probability that molecular features belong to a particular volunteer. Source tracking data based on volunteer 1 is shown in (**a,e**). Source tracking based on volunteer 2 is (**b,f**). Source tracking based on volunteer 3 is (**c,g**). Source tracking based on volunteer 4 are (**d,h**).

Determining who touched what. In addition to individual molecules and microbes being mapped in 3D space and also each inhabitant's "biome" possessing individual uniqueness, the interaction of inhabitants with their environments was also mapped (Fig. 4). SourceTracker[9] is a Bayesian analysis that has previously been used to detect contamination in high-throughput metagenomic studies[17]. SourceTracker was used in this study to delineate the molecular interaction between each volunteer and the built environment. SourceTracker successfully correlated molecular profiles (MS accuracy = 0.742857142857, 16S accuracy = 0.761904761905) collected

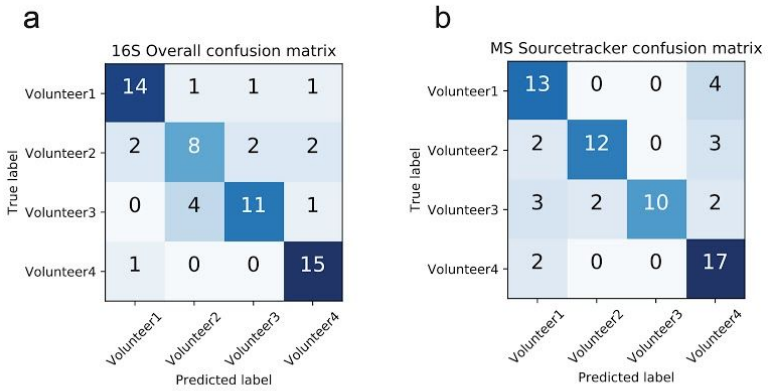


Figure 5. SourceTracker V.1 confusion matrices of metabolomic and microbial accuracy. (**a**) Using 16S input data in SourceTracker, 14 out of 17 total objects were correctly identified to belong to volunteer 1, 8 out of 14 objects were correctly identified to belong to volunteer 2, 11 out of 16 objects were correctly identified to belong to volunteer 3, and 15 out of 16 objects were correctly identified to belong to volunteer 4. (**b**) Using MS input data in SourceTracker, 13 out of 17 objects were correctly identified to belong to volunteer 1, 12 out of 17 objects were correctly identified to belong to volunteer 2, 10 out of 17 objects were correctly identified to belong to volunteer 3, and 17 out of 19 objects were correctly identified to belong to volunteer 4. Differences between total number of objects between volunteers were a result of each volunteer owned different amounts of objects. Differences between 16S and MS total amount of objects is a result of not having all of the complementary MS samples and 16S samples run experimentally.

from the office space to those collected from participants (Fig. 5a,b). We show that chemistries from sample sites (table, desk and floor) within the office are highly correlated to the main inhabitant of the office space volunteer 3 (Fig. 4c,g). The chemical and microbial distribution of volunteers 1, 2, and 4 are limited to sites closer in spatial proximity to where the volunteers were physically sitting or objects they frequently interacted with that they brought into the room, i.e. phones, computers, and chair armrests (Fig. 4a,b,d–f,h).

Discussion

Over the last decade, advances in mass spectrometry and DNA sequencing has made the technology cheaper and more accessible[21]. Although both have been widely used in forensics for some time[22–25], it has yet to be used in the context of a global microbial and molecular inventory. Despite that microbial and untargeted metabolomics analysis for routine crime scene reconstruction forensics is still premature, modern microbial DNA and untargeted mass spectrometry technologies can provide many clues that would otherwise remain hidden from sight. These clues can be used in addition to existing forensic techniques in the reconstruction of a crime scene or potentially narrow down candidate suspects. Furthermore, this proof-of-concept experiment explored the value of combining molecular networking, PLS, Source tracking and 3D cartography and its potential application towards future forensics investigations.

Molecular networking not only revealed the origins of individual molecules, but also revealed an immense chemical overlap between inhabitants and the office space. Additionally, molecular networking was able to provide greater insight into where environmental molecules originate even if they didn't match directly to an individual. As shown in Fig. 1f, nobiletin was found on the computer of volunteer 4, as well as the phones of volunteer 2 and 3. Since this molecule was not found on any individual, it was difficult to establish with confidence the ownership of this molecular feature. Figure 1g illustrates that although nobiletin was not matched specifically to an individual, it was matched to the lip balm personal care product of volunteer 4. This helped investigators hypothesize that the origin of the nobiletin is from volunteer 4 and was somehow shared to the personal devices of volunteer 2 and 3, which most likely occurred through direct contact. Without seeing direct overlap of chemicals form a person to a place, objects from suspects provided added evidence to establish origin of molecules. This illustrates the value in including non-human evidence or personal belongings that may be found outside of the investigation site. Discovering this relationship is important when identifying an individual's chemical contribution to the surrounding environment.

More often than not, inhabitants of a place may not be present at the time of sample collection. Suspect formulation commonly rely on mugshot[26] or fingerprinting[27] while more modern approaches include technologies such as biometrics[28,29], all which in isolation present clear ambiguity and biases[30,31]. This investigation supports the inclusion of lifestyle traces as a means to establish suspects within a given environment. As seen in Fig. 1, molecular annotation can be used to assign lifestyle clues of environmental inhabitants and can provide investigators with information of where inhabitants (or suspects) have been or where they may be going.

The widely distributed molecules like laureth sulfate (common household detergent) (1c), theophylline (caffeine derivative) (1g), avobenzone (sunscreen ingredient) (1a) (2D-erythro-sphingosine (eukaryotic lipid) (1d)

may not be an effective molecular feature to formulate suspects in the office, but DEET (insect repellent) (1b) and Norvasc™ (medication) (1e) may be useful since they are specific to volunteers 3 and 1 respectively. These molecules are also found in close proximity in the built environment to these volunteers (Fig. 1) and may provide supplementary clues towards who is in the room and where they may frequent. Furthermore, molecules like azoxystrobin (common food contaminant pesticide[32]) found on the computer, hands, and phone of volunteer 2 illustrate that even molecules that are inadvertently encountered can serve as molecular signatures in trace levels. The microbial diversity and uniqueness found in the office space provides similar insights to the molecular annotations. It was expected to see soil and human bacteria (Fig. 1i–k,m–p) present in the room and unique to each individual. Surprisingly, marine bacteria unique to volunteer 3 (Fig. 1l) was an interesting supporting clue that separates this individual from the rest of the volunteers and also strongly suggests volunteer's 3 interaction with the ocean. This could have significant implications when addressing who was present in a space, but as with any investigation, would require additional sampling from both environment and suspect.

A question that may arise when recreating a crime scene is to whether or not the human derived microbes or molecules are unique between each individual in the environment. In addition to demonstrating humans share a common space of molecular and microbial signatures, this investigation also supports that the biomes of each participant are unique and separate from each other at least in this particular instance in time. Although many molecules and microbes are shared across common space, PLS was able to illustrate that there is a subset of molecules and microbes that are unique to individuals (Fig. 2), which is supported by Indicator Value Analysis. As seen in Supplementary Table S2, the marine microbes identified on volunteer 3 may be useful as microbial fingerprints throughout the environment. Rather than attempting to identify possible suspects in a particular area based on their metabolome or microbiome, the uniqueness of each individual's biome suggests their biome may be of value when determining their interaction with the surrounding environment. Although not yet directly demonstrated experimentally, we anticipate that PLS could be used in the future to help identify individuals to their microbes and molecules.

A key goal of this investigation was to understand how potential suspects interacted with the habitat and objects therein. Mapping SourceTracker in 'ili illustrated interactions from both the microbe and molecular information (Fig. 3). Although SourceTracker accuracies hover at approximately ~75%, the successful correlation between person and objects at this scale successfully illustrate value in SourceTracker analysis. While volunteers (or potential suspects) 1, 2 and 4 interact on a molecular level in close proximity to each other. Volunteer 3 looks to interact more openly with the environment. Using primarily SourceTracker information, investigators concluded that volunteer 3 is the main inhabitant of the environment and much of the visiting inhabitants interaction is limited to their personal spaces or devices. This is reflected in the most significant associations of molecules and microbes to the floor, desk, table, and office phone. One can even gain insight into how this person uses the office as the molecules and microbes match best on the left side as this person enters the door to sit in the chair this volunteer is currently occupying. Additionally, the data suggests that volunteer 3 uses the right side of the office less frequently. SourceTracker also illustrated that much of the molecules and microbes belonging to volunteers 1, 2, and 4 are within close spatial proximity to the volunteers themselves. The localized distribution of these volunteer's molecules and microbes reinforce the fact that they don't interact with the environment often or at least as much as volunteer 3.

This study demonstrated an integrative approach of 3D cartography with mass spectrometric and microbial inventories to distinguish people within a group and also from their interactions with the built environment. Additionally, molecular and microbial annotations were used to predict behaviors and visited environments that individuals may experience outside of the investigated environment. Given the task to link, distinguish and predict how individuals interact with their environment, this approach offers alternative methods to traditional forensic techniques.

As with any investigation, scientific or judicial, controlling for error is of paramount importance. Although this proof-of-concept study provides novel insights into emerging molecular forensics capabilities, there are ethical, social and legal implications that surround such technologies. The scientific forensic community will have to tackle these societal and privacy challenges in the next few decades as the technology matures. Further developing the reliability and statistical accuracy of human-environmental interaction would greatly strengthen its merit in a courtroom, but in its current implementation it may help an investigator reconstruct what transpired in a room. It is important to note that the current state of this technology cannot and should not replace traditional forensics approaches. Additionally, the authors of this study hope that this proof-of-concept work will launch new research that will test temporal limits (how long do signatures last) and scalability (including multiple rooms, large number of suspects, etc.) of the approach introduced here.

Methods

Location selection. Office space is one of the most common human-occupied built environments in the United States, and was therefore selected as the primary site for this proof-of-principle study.

Volunteers/IRB protocol #. Four volunteers were recruited to participate in this study. All Individuals signed a written informed consent and approved sample collection form their hands and personal objects in accordance with sampling procedures approved by the University of California San Diego Institutional Review Board protocol (IRB number **130537**×).

Sampling. Volunteers were sampled 39 times across their body by taking individually pre-soaked EtOH/H20 cotton swabs and pressing to skin. Subsequently individually pre-soaked EtOH/H20 cotton swabs were pressed and rubbed against surfaces of the built environment. Each site on volunteer and sampling location was sampled twice for metabolite profiling and 16S microbial analysis. Sampling was performed at a 2 × 2 cm area of the head, forehead, cheek, nose, chin, bicep, forearm, hand, upper leg, lower leg, back, and bottom of shoe for each

volunteer. Various 2 × 2 cm sample sites across the built-environment were also taken including chairs, shelves, desks, computers, mouses, cell phones, personal devices and phones. Pre-moistened cotton swabs were used at each site in 50:50 ethanol/water for MS analysis or 50 mM Tris pH 7.6, 1 mM EDTA, and 0.5% Tween 20 for nucleic acid analysis. Sample locations were noted, and swabs were placed in a deep-well 2-mL polypropylene 96-well micro titer plate.

Extraction. DNA extraction is performed to capture microbial DNA from each sample that can be subjected to sequencing. The absorbed material was extracted in 500 μL of 50:50 ethanol/water (for mass spectrometric analysis) or Tris-EDTA buffer (50 mM Tris pH 7.6, 1 mM EDTA, and 0.5% Tween 20) for bacterial DNA extraction. After DNA extraction, amplification and sequencing according to the 16S rRNA amplicon protocol of the Earth Microbiome Project v.4.13 was performed[16]. DNA was extracted, amplified and sequenced at UC San Diego according to the 16S rRNA amplicon protocol of the Earth Microbiome Project v.4.13[16].

UPLC-MS/MS. At this stage of the methods, we separated molecules using chromatography and detected the ions of the molecules using a mass spectrometer. Processed office/volunteer extracts (5 uL) were subjected to UHPLC chromatographic separation using an UltiMate 3000 UHPLC system (Thermo Scientific), controlled by Chromeleon software (Thermo Scientific). Chromatographic separation was achieved using an 1.7 micron C18 (50 × 2.1 mm) Kinetex UHPLC column (Phenomenex) at 40 °C, using a flow rate of 0.5 mL/min. A linear gradient was used for the separation: 0–0.5 min 5% B, 0.5–8 min 5%B-99%B, 8–9 min 99%B, 9.01–10 min 1%B, 10–10.5 min 5–99% B, 11–11.5 min 99%B-1%B, 12–12.5 min 1% B where solvent A is water 0.1% formic acid (v/v) and solvent B is acetonitrile with 0.1% formic acid (v/v). Column eluent was introduced directly into a Bruker Daltonics maXis Impact quadrupole-time-of-flight mass spectrometer equipped with an Apollo II electrospray ionization source and controlled via otofControl v3.4 (build 16) and Hystar v3.2 software packages (Bruker Daltonics). The maXis instrument was first externally calibrated using ESI-L Low Concentration Tuning Mix (Agilent Technologies) prior to initiation of the sequence of samples, and hexakis (1H,1H,2H-difluoroethoxy) phosphazene (Synquest Laboratories), m/z 622.0295089613, was continuously introduced as an internal calibrant (lock mass) during the entirety of each LC/MS run. Data was collected in positive ion mode, scanning from 80–2000 m/z. Instrument source parameters were set as follows: nebulizer gas (Nitrogen) pressure, 2 Bar; Capillary voltage, 4,500 V; ion source temperature, 200 °C; dry gas flow, 9 L/min. MS1 spectral acquisition rate was set at 3 Hz and MS/MS acquisition rate was variable (5–10 Hz) depending on precursor intensity. Data-dependent MS/MS acquisition was programmed to the top five most intense precursors per MS^1 scan and any precursor was actively excluded for 1 minute after being fragmented twice. Each MS/MS scan acquired was the average of 4 collision energies, paired optimally with specific collision RF (or "ion cooler RF") voltages and transfer times in order to maximize the qualitative structural information from each precursor. The auto-acquisition of MS/MS spectra was carried according to specific settings. Precursor *m/z* 100, 300, 500, 1000 was selected; with an isolation width of 2, 4, 6, 8; with a base collision energy (eV) of 10, 25, 30, 50; a sampled collision energy of (5, 10, 15, 20), (12.5, 25, 37.5, 48), (15, 30, 45, 60), (25, 40, 75, 100); a collision RF (Vpp) of 250, 500, 1000, 1500 for all precursors; and a transfer time (μsec) of 50, 75, 100, 150 for all samples was used respectively. Data is publically available at http://gnps.ucsd.edu under accession number MSV000079181.

Feature finding (MS data). Feature finding is used to identify and distinguish detectible ions (molecules) from the complex sample mixture. Open-source OpenMS/TOPP software[19] was used to identify molecular features from processed featureXML format. Unique features, represented by MS1 molecular mass, were identified based on detected precursor mass and retention time. An intensity table relating abundances of molecular feature per sample site was obtained automatically from samples. Feature detection parameters include: Threshold 0.95, *m/z* tolerance 10 ppm, and retention time tolerance 10 sec.

Sequencing. DNA extraction and V4 paired end sequencing[15,16] from samples were performed according to the Earth Microbiome Project Protocols[16] and sequenced using an Illumina MiSeq (La Jolla, CA) to determine what bacteria are present at each sample site. Although V4 or any other amplification primer set has biases toward amplifying different organisms from humans[33–35], including skin[36], but also environmental microbes[16], we and others have shown that V4 primers are sufficient to be used to trace them to people[36–40] however as sequencing becomes cheaper performing the same task with metagenomes will increase the resolution with which the work outlined in this paper might be performed[41]. Processed tables can be found in https://qiita.ucsd.edu/study/description/10244, in addition sequences can be found in EBI under accession number ERP020615. Amplicon sequences were demultiplexed and quality controlled using the defaults as provided by QIIME 1.9.1. The primary sOTU table was generated using Qiita which uses deblurring[42].

QIIME. QIIME is an open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data[43]. QIIME 1.9.1 was used to compute Hellinger distances on the MS and 16S datasets. The distance matrices were partitioned by sample type. Mantel tests were performed between the MS and 16S data controlling for the sample type. Principal coordinates were computed per sample type. Taxonomy of 16S sequences were assigned using the SortMeRNA method using Greengenes reference database.

Microbiome/Metabolite correlations. Partial Least Squares (PLS) was used to correlate microbes and metabolites. This was performed using the PLSSVD function in scikit-learn, which calculates an SVD on the covariance matrix between log transformed metabolite and microbial abundances. A pseudocount was added to both datasets to avoid taking logs of zero. The loadings calculated from PLS were visualized using a biplot, where the points represent metabolites, and the arrows represent microbes. The angles between the arrow approximate the correlation between the corresponding microbes, and the distance between the points approximate the correlation

between the corresponding metabolites[44] package identify which microbes and metabolites are likely to be uniquely associated to specific volunteers. Points and arrows were colored by the volunteers that the corresponding metabolites and microbes were most likely associated with, with a p-value less than 0.001 and 1000 permutations.

SourceTracker. A computational tool initially designed to predict sources and proportions of microbial contaminations in functional metagenomics studies was used. Bayesian statistics were calculated using SourceTracker V.1[9] to determine the proportion of microbes and metabolites from the volunteers that were present in the objects. A confusion matrix was created to describe the performance of the SourceTracker classification model. True ownership of objects were known and tested by SourceTracker to achieve an accuracy value.

GNPS dereplication. Global Natural Products Social Molecular Networking[7] (http://gnps.ucsd.edu) dereplication was used to identify MS/MS spectral matches of molecular features in each sample set. Parameters for creating molecular network include a parent mass tolerance = 1.0 Da, Min matched Peak = 5, Ion Tolerance = 0.5, Score Threshold = 0.60 with default advance and filter search options (http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0a0a825f41da4781a374cfb609283097). Using these parameters, 201,563 MS/MS spectra were merged into 17565 nodes. Raw data were uploaded to http://gnps.ucsd.edu and cataloged with a massive ID of MSV000079181.

GNPS molecular networking. Global Natural Products Social Molecular Networking[7] was used to identify MS/MS spectral matches of molecular features in each sample set. Parameters for creating molecular network include a parent mass tolerance = 0.1 Da, Min matched Peak = 6, Ion Tolerance = 0.5, Score Threshold = 0.80 with default advance and filter search options (http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0a0a825f41da4781a374cfb609283097). Using these parameters, 301,113 MS/MS spectra were merged into 21,732 nodes. Raw data were uploaded to http://gnps.ucsd.edu and cataloged with a massive ID of MSV000079181. Networking parameters can be found with GNPS ID: http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=afed787019764b829ca7bf406500668f.

3D Cartography. Molecular feature visualization was performed to visualize microbiome and metabolome feautres on a 3D office model. The open-source web application 'iliwas used by incorporating feature finding tables and STereoLithography (.stl) file (Supplementary material). For creating the 3D model, we used a 3D scanner (Structure bracket kit) attached to an iPad mini (Apple, Mountain View, CA, USA). Structure SDK application software was used to render and export the 3D model into the.stl file format.

Data Availability. Analyses notebooks: https://github.com/knightlab-analyses/office-study. QIITA File: https://qiita.ucsd.edu/study/description/10244. Mass spectrometry is deposited in GNPS: Accession number MSV000079181. GNPS Molecular networking: http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=afed787019764b829ca7bf406500668f. Cartographical snapshot link for the 3D visualization of SourceTracker results for 16S: https://ili.embl.de/?ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/Model_wall%20(1).stl; ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/16S_ILI/sourcetracker_16s_ili_mapping2.csv; ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/16S_ILI/Volunteer%201.json. Cartographical snapshot link for the 3D visualization of SourceTracker results for MS: https://ili.embl.de/?ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/Model_wall%20(1).stl; ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/MS_ILI/sourcetracker_ms_ili_mapping.csv; ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/MS_ILI/Volunteer_1.json. Cartographical snapshot link for the 3D data of molecules (MS) and bacterial taxa (16S): https://ili.embl.de/?ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/Model_wall%20(1).stl; ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–22_ckapono_11de23dd/other/ili_complete_fig_3.csv;ftp://massive.ucsd.edu/MSV000079181/updates/2017–08–16_ckapono_f9095c5e/other/3D%20office%20Source%20Tracker%20supplemental%20files/MS_ILI/Volunteer_1.json. Sequence data EBI accession number: ERP020615.

References

1. Petras, D. *et al.* Mass spectrometry-based visualization of molecules associated with human habitats. *Anal. Chem.* **88**(22), 10775–10784 (2016).
2. Humphries, C. Indoor ecosystems. *Science.* **335**(6069), 648–650 (2012).
3. Dunn, R. R., Fierer, N., Henley, J. B., Leff, J. W. & Menninger, H. L. Home life: factors structuring the bacterial diversity found within and between homes. *PLoS ONE* **8**(5), e64133 (2013).
4. Bouslimani, A. *et al.* Lifestyle chemistries from phones for individual profiling. *Proc Natl Acad Sci.* **113**(48), E7645–E7654 (2016).
5. Cuzuel, V. *et al.* Origin, Analytical Characterization, and Use of Human Odor in Forensics. *J Forensic Sci.* **62**(2), 330–350 (2017).
6. Metcalf, J L. *et al.* Microbiome tools for forensic science. *TRENDS Biotechnol.* (2017).
7. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* **34**(8), 828–837 (2016).
8. Boulesteix, A.-L. & Korbinian Strimmer. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform.* **8**(1), 32–44 (2006).
9. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nature Methods.* **8**(9), 761–763 (2011).
10. Protsyuk, I. *et al.* 3D molecular cartography using LC–MS facilitated by Optimus and'ili software. *Nat. Protoc.* **13**(1), 134 (2018).
11. Quinn, R. A. *et al.* Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends Pharmacol Sci.* **38**, 2 (2017).

12. Bouslimani, A. *et al*. Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci*. **112**(17), E2120–E2129 (2015).
13. Sumner, Lloyd W. *et al*. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**(3), 211–22, (2007).
14. Rowell, F., Hudson, K. & Seviour, J. Detection of drugs and their metabolites in dusted latent fingermarks by mass spectrometry. *Analyst*. **134**(4), 701–707 (2009).
15. Caporaso, J. G. *et al*. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. **6**(8), 1621 (2012).
16. Gilbert, J. A. *et al*. The Earth Microbiome Project: meeting report of the 1st EMP meeting on sample selection and acquisition at Argonne National Laboratory October 6th 2010. *Stand Genomic Sci*. **3**(3), 249 (2010).
17. Visca, P., Seifert, H. & Kevin, J. Towner. Acinetobacter infection–an emerging threat to human health. *IUBMB life*. **63**, 12,1048–1054 (2011).
18. Erlacher, A. *et al*. Rhizobiales as functional and endosymbiontic members in the lichen symbiosis of Lobaria pulmonaria L. *Front Microbiol*. **6**, 53 (2015).
19. Sturm, M. *et al*. OpenMS–an open-source software framework for mass spectrometry. *BMC Bioinformatics*. **9**(1), 163 (2008).
20. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. **28**(24), 3211–7, (2012).
21. Aksenov, A. A. *et al*. Global chemical analysis of biology by mass spectrometry. *Nat. Rev. Chem*. **1**(7), s41570–017 (2017).
22. Takats, Z., Justin, M. W. & Cooks, R. G. Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. *IJMS*. **40**(10), 1261–1275 (2005).
23. Benson, S. *et al*. Forensic applications of isotope ratio mass spectrometry—a review. *Forensic Sci. Int*. **157**(1), 1–22 (2006).
24. Wilson, M. R. *et al*. Guidelines for the use of mitochondrial DNA sequencing in forensic science. *Crime Lab Digest*. **20**(4), 68–77 (1993).
25. Wilson, M. R. *et al*. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Legal Med*. **108**(2), 68–74 (1995).
26. Lindsay, R. C. L. *et al*. Using mug shots to find suspects. *J. Appl. Psychol*. **79**(1), 121 (1994).
27. Buchanan, M. V., Asano, K. & Bohanon, A. Chemical characterization of fingerprints from adults and children. *J Forensic Investigation*. **2941**, 89–96 (1997).
28. Maltoni, D., Maio, D., Jain, A. K. & Prabhakar, S. *Handbook of fingerprint recognition*. 57–95 (Springer, 2009).
29. Jain, A. K., Ross, A. & Prabhakar, S. An introduction to biometric recognition. *IEEE Trans*. **14**(1), 4–20 (2004).
30. Broeders, A. P. A. Of earprints, fingerprints, scent dogs, cot deaths and cognitive contamination—a brief look at the present state of play in the forensic arena. *Forensic Sci. Int*. **159**(2), 148–157 (2006).
31. Brown, E., Deffenbacher, K. & Sturgill, W. Memory for faces and the circumstances of encounter. *J. Appl. Psychol*. **62**(3), 311 (1977).
32. Kanetis, L., Förster, H. & James, E. Adaskaveg. Comparative efficacy of the new postharvest fungicides azoxystrobin, fludioxonil, and pyrimethanil for managing citrus green mold. *Plant Dis*. **91**(11), 1502–1511 (2007).
33. Shepherd, M. L. *et al*. Characterization of the fecal bacteria communities of forage-fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol Lett*. **326**(1), 62–68 (2012).
34. Kuczynski, J. *et al*. Experimental and analytical tools for studying the human microbiome. *Nature Rev Genet*. **13**(1), 47–58 (2012).
35. Fadrosh, D. W. *et al*. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. **2**(1), 6 (2014).
36. Castelino, M. *et al*. Optimisation of methods for bacterial skin microbiome investigation: primer selection and comparison of the 454 versus MiSeq platform. *BMC Mol Biol*. **17**(1), 23 (2017).
37. Huttenhower, C. *et al*. Structure, function and diversity of the healthy human microbiome. *Nature* **486**(7402), 207 (2012).
38. Clemente, J. C. *et al*. The microbiome of uncontacted Amerindians. *Sci Adv*. **1**(3), e1500183 (2015).
39. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res*. **19**(7), 1141–1152 (2009).
40. Lax, S. *et al*. Forensic analysis of the microbiome of phones and shoes. *Microbiome*. **3**(1), 21 (2015).
41. Franzosa, E. A. *et al*. Identifying personal microbiomes using metagenomic codes. *Proc Natl Acad Sci*. **112**(22), E2930–E2938 (2015).
42. Amir, A. *et al*. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*. **2**(2), e00191–16 (2017).
43. Caporaso, J. G. *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. **7**(5), 335–336 (2010).
44. Roberts, D.W. and Roberts, M.D.W. Ordination and Multivariate Analysis for Ecology. Bozeman, Montana. http://ecology.msu.montana.edu/labdsv/R (2016).

Acknowledgements

We thank all volunteers who were recruited in this study for their participation. This work was supported by National Institute of Justice Award 2015-DN-BX-K047 and the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 634402. This work was partially supported by US National Institutes of Health (NIH) Grant 5P41GM103484–07 and by the Alfred P. Sloan Foundation. We further acknowledge NIH Grant GMS10RR029121 and Bruker for the shared instrumentation infrastructure that enabled this work.

Author Contributions

Pieter Dorrestein created the idea of the project, provided guidance of the analysis and interpretation of the data, helped to create the 3D images, performed molecular networking and metabolomic analysis and helped to write the paper. Rob Knight provided guidance on the overall analysis and interpretation of the data, interpreted microbial and metabolome correlations to volunteers and helped to write the paper. Cliff Kapono provided guidance on the analysis and interpretation of the data, collected the mass spectrometry data, performed the statistical analysis, created the 3D images, performed molecular networking and analysis of the networks and wrote the paper. James Morton provided guidance on the analysis and interpretation of the data, performed statistical analysis, created 3D images, interpreted microbial and metabolome correlations to volunteers and helped to write the paper. Kayla Orlinsky provided statistical guidance and helped to create the 3D images. Stefan Janssen provided guidance on analysis and interpretation of the data and helped write the paper. Qiyun Zhu provided guidance on analysis and interpretation of the data. Amina Bouslamani helped to collect samples and write the paper. Alexey Melnik helped to collect samples. Tal Luzzatto Knaan helped to collect samples. Neha Garg provided guidance and analysis of data. Theodore Alexandrov created 'ili and helped write the paper. Ivan Protsyuk developed 'ili. Larry Smarr provided guidance and analysis of data. Yoshiki Vázquez-Baeza provided statistical guidance and analysis of the data.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-21541-4.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

Online Supporting Material for: "Creating a 3D microbial and chemical snapshot of a human habitat"

Authors: Clifford A. Kapono[1], James T. Morton[3,7], Amina Bouslimani[2], Alexey V. Melnik[2], Kayla Orlinsky[3] Tal Luzzatto Knaan[2], Neha Garg[2], Yoshiki Vázquez-Baeza[3], Ivan Protsyuk[4], Stefan Janssen[7], Qiyun Zhu[7], Theodore Alexandrov[2,4], Larry Smarr[5,6], Rob Knight[2,6,7]*, Pieter C. Dorrestein[2,6]**

Author Affiliations:
1. Department of Chemistry, University of California San Diego, La Jolla, CA, USA
2. Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA, USA
3. Department of Computer of Science and Engineering, University of California San Diego, La Jolla, CA, USA
4. Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
5. California Institute for Telecommunications and Information Technology, University of California San Diego, La Jolla, CA, USA
6. Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA
7. Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

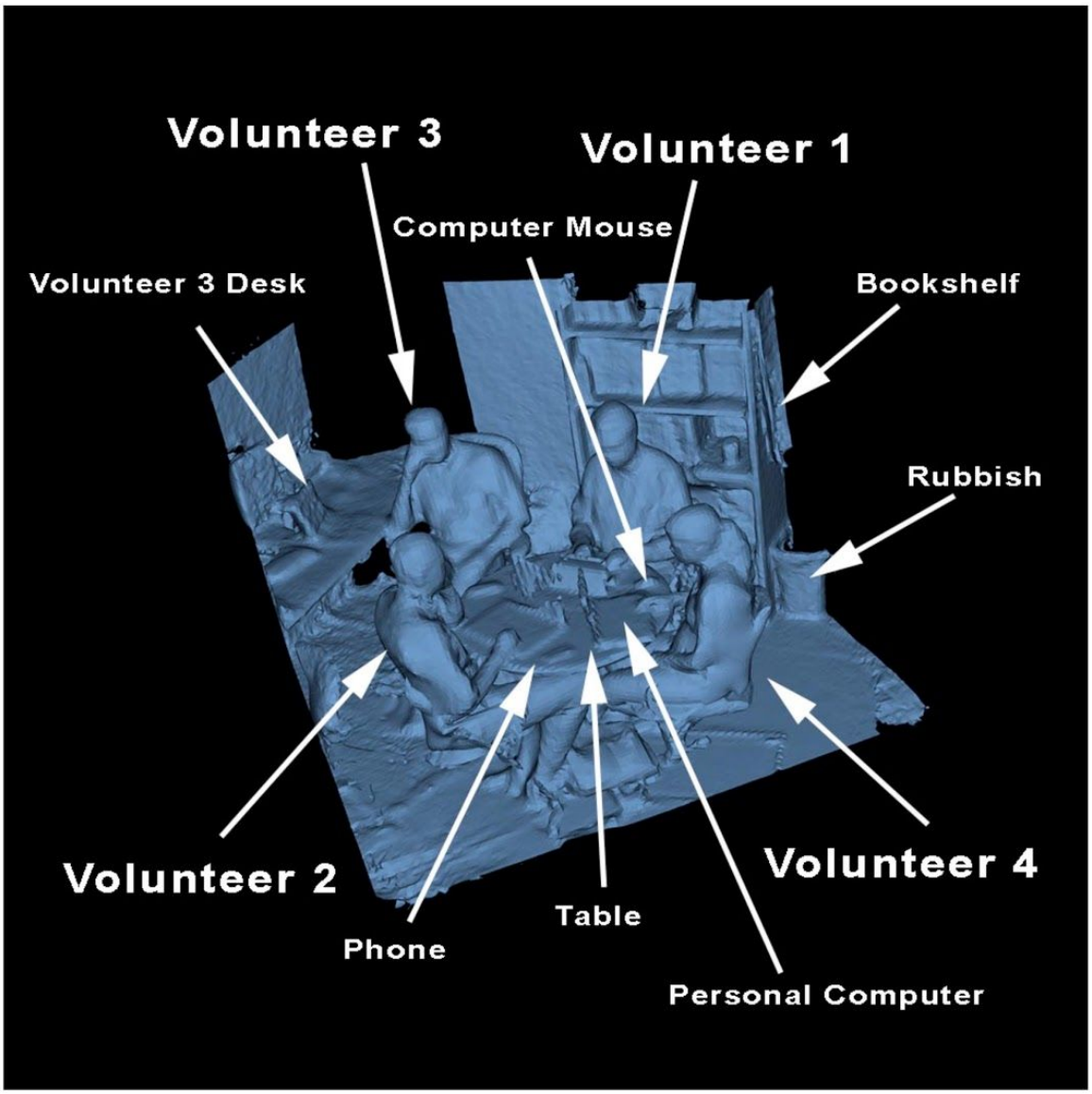
* To whom correspondence should be addressed regarding sequencing and statistics rknight@ucsd.edu ** to whom correspondence should be addressed regarding the project, 3D analysis and mass spectrometry pdorrestein@ucsd.edu.

In this online support material, we show (1) a 3D cartography map of the office environment and its inhabitants, (2) SortMeRNA[1]-assigned taxonomy microbiome correlations in respects to each inhabitant of the office space (3) GNPS[2]-assigned molecular annotation correlations in respects to each inhabitant of the office space and (4) SortMeRNA[1]-assigned taxonomy microbe correlations to volunteer 3.

1. Kopylova, Evguenia, Laurent Noé, and Hélène Touzet. "SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data." Bioinformatics 28.24 (2012): 3211-3217.

2. Wang, Mingxun, *et al*. "Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking." *Nature Biotechnology* 34.8, 828-837 (2016).

Supplementary Figure S1. 3D map of an office built environment. Four occupants are sitting in chairs at a table with their personal computers and cellular phones. They are surrounded by a bookshelf, and the desk of volunteer 3.



Supplementary Table S1. Microbiome Volunteer Correlations

Volunteer	Sequence	SortMeRNA-assigned taxonomy	Indicator value	p-value
1	TACGAAGGGGGCTA GCGTTGTTCGGAATT ACTGGGCGTAAAGC GCACGTAGGCGGAC ATTTAAGTCAGGGGT GAAATCCCGGGGCT CAACCTCGGAACTG	*Agrobacterium*	0.90	0.001
1	TACGAAGGGGGCTA GCGTTGTTCGGAATC ACTGGGCGTAAAGC GCACGTAGGCGGAC CATTAAGTCAGGGGT GAAAGCCTGGAGCT CAACTCCAGAACTG	*Rhizobiales*	0.83	0.001
1	TACGTAGGGCGCGA GCGTTGTCCGGAATT ATTGGGCGTAAAGA GCTTGTAGGCGGTT GGTCGCGTCTGCTG TGAAAGGCTGGGGC TTAACCCTGGTTTTG	*Rothia dentocariosa*	0.79	0.001
2	TACGTAGGTGGCAA GCGTTATCCGGAATT ATTGGGCGTAAAGC GCGCGTAGGCGGTT TCTTAAGTCTGATGT GAAAGCCCACGGCT CAACCGTGGAGGGT	*Staphylococcus*	0.91	0.001
2	TACGTAGGGTGCGA GCGTTGTCCGGAATT ACTGGGCGTAAAGA GCTCGTAGGTGGTTT GTCGCGTCGTCTGT GAAATCCCGGGGCT TAACTTCGGGCGTG	*Corynebacterium*	0.87	0.001
2	TACGTAGGGTGCGA GCGTTGTCCGGAATT ACTGGGCGTAAAGA GCTCGTAGGTGGTTT GTCGCGTCGTTTGT	*Corynebacterium*	0.85	0.001

	GTAATACCGCAGCTT AACTGCGGGGTTG			
3	TACGTAGGGTGCGA GCGTTGTCCGGAATT ACTGGGCGTAAAGA GCTCGTAGGTGGTTT GTCGCGTCGTTTGT GGAATACCGCAGCT TAACTGTGGGGTTG	*Corynebacterium*	0.94	0.001
3	TACGAAGGTCCCAA GCGTTGTTCGGAATA ACTGGGCGTAAAGC GTGTGTAGGCTGCG CGGAAAGTCAAATGT GAAAGCCAAGGGCT CAACCCTTGAACTG	*Verrucomicrobiaceae*	0.78	0.001
3	TACGGAGGGTGCGA GCGTTATCCGGAATC ACTGGGCGTAAAGG GCGTGTAGGCGGGA CGTTAAGTCTGGTTT TAAAGACCGCAGCT CAACTGCGGGAGTG	*Deinococcus*	0.75	0.001
4	TACGGAGGGTGCGA GCGTTATCCGGAATC ACTGGGCGTAAAGG GCGTGTAGGCGGGA CGTTAAGTCTGGTTT TAAAGACCGCAGCT CAACTGCGGGAGTG	*Deinococcus*	0.82	0.001
4	TACGAAGGGTGCAA GCGTTACTCGGAATT ACTGGGCGTAAAGC GTGCGTAGGTGGTC GTTTAAGTCCGTTGT GAAAGCCCTGGGCT CAACCTGGGAACTG	*Xanthomonadaceae*	0.79	0.001
4	TACGTAGGGTGCGA GCGTTGTCCGGAATT ACTGGGCGTAAAGA GCTCGTAGGTGGTTT GTCGCGTCGTTTGT GTAAGTCCACAGCTT AACTGTGGGACTG	*Corynebacterium*	0.77	0.001

Supplementary Table S2. Metabolomics Volunteer Correlations

Volunteer	Molecular Weight (*m/z*)	Retention Time (sec)	GNPS Search Result	Indicator value	p-value
1	388.39	444-584	No Matches	0.86	0.001
1	282.27	448-583	No Matches	0.81	0.001
1	192.08	193-203	No Matches	0.78	0.001
2	374.33	336-361	No Matches	0.86	0.001
2	282.27	388-413	No Matches	0.81	0.001
2	192.08	402-445	No Matches	0.78	0.001
3	664.51	577-601	No Matches	0.89	0.001
3	669.46	574-599	No Matches	0.87	0.001
3	655.99	566-601	No Matches	0.86	0.001
4	368.42	465-603	No Matches	0.96	0.001
4	340.39	444-599	No Matches	0.90	0.001
4	312.35	410-575	No Matches	0.84	0.001

Supplementary Table S3. Marine Bacteria present on Volunteer 3

Sequence	Indicator value	*p*-value	SortMeRNA-assigned taxonomy
TACGGAGGGTGC GAGCGTTAATCGG AATTACTGGGCGT AAAGCGCGTGTAG GTGGTTAATTAAG TCAGATGTGAAAG CCCAGGGCTCAAC CCTGGAACTG	0.181818182	0.001	*Psychromonas*
TACAGAGGGTGCA AGCGTTAATCGGA ATTACTGGGCGTA AAGCGCGCGTAG GTGGTTTGTTAAG TCTGATGTGAAAT	0.267170702	0.001	*Marinomonas*

CCCAGGGCTCAAC CTTGGAATGG			
TACGGAGGGTGC GAGCGTTAATCGG AATTACTGGGCGT AAAGCGCGCGTA GGTGGTTAGTTAA GTCAGATGTGAAA TCCCAGGGCTCAA CCTTGGAACTG	0.27090713	0.001	*Psychromonas*
TACGGAGGGTGC GAGCGTTAATCGG AATTACTGGGCGT AAAGCGCACGCA GGCGGTTTGTTAA GCTAGATGTGAAA GCCCCGGGCTCA ACCTGGGACGGT	0.333333333	0.001	*Alteromonas*
TACGGAGGGTGC GAGCGTTAATCGG AATTACTGGGCGT AAAGCGCATGCAG GTGGTTCATTAAG TCAGATGTGAAAG CCCGGGGCTCAA CCTCGGAACTG	0.388894577	0.001	Vibrionaceae
TACGGAGGGTGC GAGCGTTAATCGG AATTACTGGGCGT AAAGCGTACGCAG GCGGTTTGTTAAG CGAGATGTGAAAG CCCCGGGCTCAA CCTGGGAACTG	0.450119335	0.001	*Pseudoalteromonas*
TACGGAGGGTGC GAGCGTTAATCGG AATTACTGGGCGT AAAGCGCGCGTA GGCGGTTAATTAA GTCAGATGTGAAA TCCCAGGGCTCAA CCTTGGAACTG	0.555962764	0.001	*Psychromonas*

2.3 Acknowledgements

Chapter 2, in full, is a reprint of materials as it appears in “Creating a 3D microbial and chemical snapshot of a human habitat” in *Scientific Reports*, 2018, Kapono, C. A., Morton, J. T., Bouslimani, A., Melnik, A. V., Orlinsky, K., Knaan, T. L, Garg, N., Vazquez-Baeza, Y., Protsyuk, I., Janssen, S., Zhu, Q., Alexandrov, T., Smarr, L., Knight, R., Dorrestein, P.C.. The dissertation author was the primary investigator and author of this manuscript.

2.4 References

1. Cohen JE. Human population: the next half century. Science. 2003 Nov 14;302(5648):1172–5.
2. Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM. Architectural design influences the diversity and structure of the built environment microbiome. ISME J. 2012 Aug;6(8):1469–79.
3. Brooks B, Firek BA, Miller CS, Sharon I, Thomas BC, Baker R. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. Microbiome. 2014 Jan 28;2(1):1.
4. Kelley ST, Gilbert JA. Studying the microbiology of the indoor environment. Genome Biol. 2013 Feb 28;14(2):202.
5. Nakamura A, Osonoi T, Terauchi Y. Relationship between urinary sodium excretion and pioglitazone-induced edema. J Diabetes Investig. 2010 Oct 19;1(5):208–11.
6. Dunn RR, Fierer N, Henley JB, Leff JW, Menninger HL. Home life: factors structuring the bacterial diversity found within and between homes. PLoS One. 2013 May 22;8(5):e64133.
7. Flores GE, Bates ST, Caporaso JG, Lauber CL, Leff JW, Knight R. Diversity, distribution and sources of bacteria in residential kitchens. Environ Microbiol. 2013 Feb;15(2):588–96.
8. Angenent LT, Kelley ST, St Amand A, Pace NR, Hernandez MT. Molecular identification of potential pathogens in water and air of a hospital therapy pool. Proc

Natl Acad Sci U S A. 2005 Mar 29;102(13):4860–5.

9. Poza M, Gayoso C, Gómez MJ, Rumbo-Feal S, Tomás M, Aranda J. Exploring bacterial diversity in hospital environments by GS-FLX Titanium pyrosequencing. PLoS One. 2012 Aug 29;7(8):e44105.

10. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014 Aug 29;345(6200):1048–52.

11. Sugita T, Yamazaki T, Makimura K, Cho O, Yamada S, Ohshima H. Comprehensive analysis of the skin fungal microbiota of astronauts during a half-year stay at the International Space Station. Med Mycol. 2016 Mar;54(3):232–9.

12. Venkateswaran K, Vaishampayan P, Cisneros J, Pierson DL, Rogers SO, Perry J. International Space Station environmental microbiome - microbial inventories of ISS filter debris. Appl Microbiol Biotechnol. 2014 Apr 4;98(14):6453–66.

13. Moissl-Eichinger C, Cockell C, Rettberg P. Venturing into new realms? Microorganisms in space. FEMS Microbiol Rev. 2016 Sep;40(5):722–37.

14. Nordahl Petersen T, Rasmussen S, Hasman H, Carøe C, Bælum J, Schultz AC. Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance. Sci Rep. 2015 Jul 10;5:11444.

15. Flores GE, Bates ST, Knights D, Lauber CL, Stombaugh J, Knight R. Microbial biogeography of public restroom surfaces. PLoS One. 2011 Nov 23;6(11):e28132.

16. Malnick S, Melzer E. Human microbiome: From the bathroom to the bedside. World J Gastrointest Pathophysiol. 2015 Aug 15;6(3):79–85.

17. Shaughnessy MK, Bobr A, Kuskowski MA, Johnston BD, Sadowsky MJ, Khoruts A. Environmental Contamination in Households of Patients with Recurrent Clostridium difficile Infection. Appl Environ Microbiol. 2016 May;82(9):2686–92.

18. Pei XQ, Song M, Guo M, Mo FF, Shen XY. Concentration and risk assessment of phthalates present in indoor air from newly decorated apartments. Atmos Environ. 2013;68:17–23.

19. Hewitt KM, Gerba CP, Maxwell SL, Kelley ST. Office space bacterial abundance and diversity in three metropolitan areas. PLoS One. 2012 May 30;7(5):e37849.

20. Meadow JF, Altrichter AE, Kembel SW, Moriyama M, O'Connor TK, Womack AM. Bacterial communities on classroom surfaces vary with human contact. Microbiome. 2014 Mar 7;2(1):7.

21. Weschler CJ, Nazaroff WW. Semivolatile organic compounds in indoor

environments. Atmos Environ. 2008;42(40):9018–40.

22. Shea KM, Committee on Environmental Health. Pediatric Exposure and Potential Toxicity of Phthalate Plasticizers. Pediatrics. 2003;111(6):1467–74.

23. Boor BE, Liang Y, Crain NE, Järnström H, Novoselac A, Xu Y. Identification of Phthalate and Alternative Plasticizers, Flame Retardants, and Unreacted Isocyanates in Infant Crib Mattress Covers and Foam. Environmental Science & Technology Letters. 2015;2(4):89–94.

24. Geens T, Roosens L, Neels H, Covaci A. Assessment of human exposure to Bisphenol-A, Triclosan and Tetrabromobisphenol-A through indoor dust intake in Belgium. Chemosphere. 2009 Aug;76(6):755–60.

25. Loganathan SN, Kannan K. Occurrence of bisphenol A in indoor dust from two locations in the eastern United States and implications for human exposures. Arch Environ Contam Toxicol. 2011 Jul;61(1):68–73.

26. Zota AR, Riederer AM, Ettinger AS, Schaider LA, Shine JP, Amarasiriwardena CJ. Associations between metals in residential environmental media and exposure biomarkers over time in infants living near a mining-impacted site. J Expo Sci Environ Epidemiol. 2015;26(5):510–9.

27. Wirth JJ, Mijal RS. Adverse Effects of Low Level Heavy Metal Exposure on Male Reproductive Function. Syst Biol Reprod Med. 2010;56(2):147–67.

28. Glorennec P, Lucas J-P, Mandin C, Le Bot B. French children's exposure to metals via ingestion of indoor dust, outdoor playground dust and soil: Contamination data. Environ Int. 2012;45:129–34.

29. Massey DD, Habil M, Taneja A. Particles in different indoor microenvironments-its implications on occupants. Build Environ. 2016;106:237–44.

30. Fanning DW, Smith JA, Rose GD. Molecular cartography of globular proteins with application to antigenic sites. Biopolymers. 1986 May;25(5):863–83.

31. Floros DJ, Petras D, Kapono CA, Melnik AV, Ling T-J, Knight R. Mass Spectrometry Based Molecular 3D-Cartography of Plant Metabolites. Front Plant Sci. 2017 Mar 29;8:429.

32. Protsyuk I, Melnik AV, Nothias L-F, Rappez L, Phapale P, Aksenov AA. 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. Nat Protoc. 2018 Jan;13(1):134–54.

33. Bouslimani A, Porto C, Rath CM, Wang M, Guo Y, Gonzalez A. Molecular cartography of the human skin surface in 3D. Proc Natl Acad Sci U S A. 2015 Apr 28;112(17):E2120–9.

34. Garg N, Wang M, Hyde E, da Silva RR, Melnik AV, Protsyuk I. Three-Dimensional Microbiome and Metabolome Cartography of a Diseased Human Lung. Cell Host Microbe. 2017 Nov 8;22(5):705–16.e4.

35. Meilaender PC. “This is My Own, My Native Land”: Immigration, National Identity, and the Problem of Preference. In: Toward a Theory of Immigration. 2001. p. 81–103.

3. Chapter 3

3.1 Introduction

3.1.1 Coral disease in Hawaii

As coral disease continues to threaten one of the most biodiverse and economically profitable ecosystems on Earth[1], scientist across the world are expeditiously looking to develop tools that better characterize the biology and ecology of corals[2]. Advancements in culture-independent microbial analysis has allowed investigators to characterize the microbial world of the coral holobiont[3,4], although fully understanding the functional roles between the coral microbiomes and disease remains challenging[5,6]. Additionally, recent advancements in LC-MS/MS molecular profiling has allowed investigators to begin characterizing coral metabolomes with increased sensitivity and reproducibility[7–10]. Structure-from-Motion (SfM)photogrammetry is another emerging technology that investigates the relationship between 3D coral structure and ecosystem processes[11–15]. While these technologies provide valuable insights for understanding coral reefs, to date, few studies have integrated these approaches to conduct multi-omic profiling across coral landscapes, especially in the context of 3D reef models.

The Waiōpae coastline is located at the southeastern tip of Hawaii Island and supports a high abundance and diversity of coral in a network of submerged pools formed by permeable basaltic substrate. Waiōpae receives a high frequency of natural and anthropogenic disturbances, both chronic and acute[16,17]. The coral disease, Growth

Anomaly (GA), is a prominent disease throughout the global oceans and exhibits a higher prevalence at Waiōpae than any other surveyed sites throughout the Hawaiian Archipelago[1616,1716] . While research has shown this disease affects critical biological functions, there is still little known about the causation and pathogenesis of GA[18].

3.1.2 3D coral reef mapping

3D reef mapping and genomic analyses have been performed independently to investigate coral communities at Waiōpae[13,18,19], yet no studies have examined metabolomic characteristics of corals at this location. Moreover, none of these technologies have been combined to conduct 3D multi-omic profiling for corals. Here we apply an integrative tool box, capable of integrating 16S sequencing and LC-MS/MS analysis with 3D SfM photogrammetry to characterize a hawaiian coral assemblage comprising of P. lobata, M. capitata, and M. flabellata species (Figure 1). This study identified key molecular features that distinguish each coral species as well as molecular features associated with GA-affected coral tissue. Analyzing the data across the reconstructed 3D coral colonies provides a new and useful perspective for determining spatial patterns in molecular profiles. Samples were collected at both day and night along spatial gradients where the coral colonies physically interacted and along gradients where tissue was affected by GA disease. This approach can be scaled-up to determine molecular underpinnings that drive biological responses and ecological characteristics in coral communities at other locations.

3.2 Methods

3.2.1 Coral tissue sampling

84 samples were taken from Waiōpae (19°29′55″N154°49′06″W), Southeast Hawaii Island in accordance with traditional hawaiian gathering protocol and State of Hawaii guidelines (DAR permit number 53929). Tissue samples were collected with a 1-cm core from *Porites lobata* (n=12), *Montipora capitata* (n=36), and *Montipora flabellata* (n=12) corals (Supplemental figure 1). The colonies in this assemblage all interacted along their borders and samples were taken along spatial gradients perpendicular to the colony borders . Samples were also collected from from *M. capitata* tissue affected by growth anomalies (n=10) and non-GA tissue (n=10) within 3-cm of the GA lesions. 1.5 mL water samples (n=4) were also taken just above coral surface. The location of each sample was recorded by photographs and video so that we could later add these locations to the coordinated in the 3D image. Each sample was subjected to molecular profiling by mass spectrometry and DNA sequencing individually. Samples were flash frozen with liquid nitrogen and shipped overnight under dry ice to the University of San Diego,California. Samples were kept in -80C until extraction.

3.2.2 3D reconstruction of coral colonies

Images, prior to sampling, were collected using methods developed specifically for reconstructing a 3D map of the coral reef habitats that were sampled[13]. One coral assemblage was identified that contained several coral that had interacting borders of different genus (*Porites/Montipora)*, same genus different species (*M. capitata/M.*

flabellata), same species (*M. capitata/M. capitata*), and diseased tissue (*M. capitata GA*). Ground control points (GCPs) were placed at the corners of the coral assemblage with known x,y,z coordinates to enable precise spatial rectification for the resulting 3D models. A calibration grid and a scale marker were placed at the border of the survey area to validate spatial accuracy. Overlapping photographs were taken from planar and oblique angles while swimming above the coral assemblage in a boustrophodonic pattern. All photos were taken with a Canon 5D Mark III digital SLR camera with a 24–70 mm lens set at 24 mm (Canon USA Inc., New York, USA) in an Ikelite housing with an 8-in. hemispheric dome port (Ikelite Underwater Systems, Indianapolis, USA). The 8-in. hemispheric dome port significantly reduces distortion due to refraction and improves the ability of the software to accurately calibrate and align the images[20,21].

3D reconstructions of the coral assemblage were rendered using Agisoft modeling software (Agisoft LLC., St. Petersburg, Russia). Images were uploaded into the PhotoScan software and camera calibration was performed using Brown's distortion model. PhotoScan aligned all overlapping images using algorithms that detect invariant features from the overlapping image sequences. The invariant features were then used to create geometrical projective matrices and determine the exact position and orientation of the camera for each sequential image (Verhoeven et al., 2012; Westoby et al., 2012). 3D geometry was constructed on the 2D image plane using the extrinsic parameters, which define the location and orientation of the camera reference frame with respect to a known world reference frame and in conjunction with the intrinsic parameters, which link the pixel coordinates of an image point with the corresponding

coordinates in the camera reference frame. Iterative bundle adjustments were used to refine the 3D coordinates of the scene geometry and minimize reprojection error in order to create a sparse 3D point cloud that accurately represented the structure of the photographed coral reef plot. Markers were digitally annotated onto each of the GCPs using the PhotoScan software, and the known x,y,z values of the GCPs were used to optimize the alignment of the photos and to ensure the resulting models were spatially rectified. After the initial point cloud was optimized, a dense point cloud was generated and used to construct a continuous mesh surface that was then triangulated and rendered with the sequential images to create textured 3D digital surface models representing the coral assemblage at Waiōpae. The 84 coordinates were annotated onto the 3D model in GeoMagic 3D annotation software[22] that represented the actual sampling sites where tissue samples were taken from each coral. Coordinates were matched with each sample file from LC-MS/MS and 16S OTU feature tables.

3.2.3 Extraction of coral metabolites and microbes

Molecular extraction was performed on the frozen samples to capture small molecules that can be subjected to LC-MS/MS. Each sample was lyophilized, weighed and subjected to 1.0 ml/mg with 70% methanol for 4 hours in 10mL glass vials[23]. Supernatant was removed and dried until LC-MS/MS data acquisition.

DNA extraction was performed on the frozen samples to capture microbial DNA from each sample that can be subjected to sequencing. The absorbed material was extracted in 500µL of 50:50 ethanol/water (for mass spectrometric analysis) or

Tris-EDTA buffer (50 mM Tris pH 7.6, 1 mM EDTA, and 0.5% Tween 20) for bacterial DNA extraction. After DNA extraction, amplification and sequencing according to the 16S rRNA amplicon protocol of the Earth Microbiome Project v.4.13 was performed[24].

3.2.4 Data acquisition from MS and 16S sampling

Molecules were separated using chromatography and detected using a mass spectrometer. Processed coral extracts (5uL) were subjected to UHPLC chromatographic separation using an UltiMate 3000 UHPLC system (Thermo Scientific), controlled by Chromeleon software (Thermo Scientific). Chromatographic separation was achieved using an 1.7 micron C18 (50 x 2.1 mm) Kinetex UHPLC column (Phenomenex) at 40°C, using a flow rate of 0.5 mL/min. A linear gradient was used for the separation: 0-0.5 min 5% B, 0.5-8 min 5%B-99%B, 8-9 min 99%B, 9.01-10 min 1%B, 10-10.5 min 5%-99% B, 11-11.5 min 99%B-1%B, 12-12.5 min 1% B where solvent A is water 0.1% formic acid (v/v) and solvent B is acetonitrile with 0.1% formic acid (v/v). Column eluent was introduced directly into a Bruker Daltonics maXis Impact quadrupole-time-of-flight mass spectrometer equipped with an Apollo II electrospray ionization source and controlled via otofControl v3.4 (build 16) and Hystar v3.2 software packages (Bruker Daltonics). The maXis instrument was first externally calibrated using ESI-L Low Concentration Tuning Mix (Agilent Technologies) prior to initiation of the sequence of samples, and hexakis (1H,1H,2H-difluoroethoxy)phosphazene (Synquest Laboratories), m/z 622.0295089613, was continuously introduced as an internal calibrant (lock mass) during the entirety of each LC/MS run. Data was collected in

positive ion mode, scanning from 80-2000 m/z. Instrument source parameters were set as follows: nebulizer gas (Nitrogen) pressure, 2 Bar; Capillary voltage, 4,500 V; ion source temperature, 200°C; dry gas flow, 9 L/min. MS1 spectral acquisition rate was set at 3Hz and MS/MS acquisition rate was variable (5-10Hz) depending on precursor intensity. Data-dependent MS/MS acquisition was programmed to the top five most intense precursors per MS^1 scan and any precursor was actively excluded for 1 minute after being fragmented twice. Each MS/MS scan acquired was the average of 4 collision energies, paired optimally with specific collision RF (or "ion cooler RF") voltages and transfer times in order to maximize the qualitative structural information from each precursor. The auto-acquisition of MS/MS spectra was carried according to specific settings. Precursor *m/z* 100, 300, 500, 1000 was selected; with an isolation width of 2, 4, 6, 8; with a base collision energy (eV) of 10, 25, 30, 50; a sampled collision energy of (5, 10, 15, 20), (12.5, 25, 37.5, 48), (15, 30, 45, 60), (25, 40, 75, 100); a collision RF (Vpp) of 250, 500, 1000, 1500 for all precursors; and a transfer time (µsec) of 50, 75, 100, 150 for all samples was used respectively. Data is publically available at massive.ucsd.edu under accession number MSV00008228.

DNA extraction and V4 paired end sequencing[24,25] from samples were performed according to the Earth Microbiome Project Protocols[24,25] and sequenced using an Illumina MiSeq (La Jolla, CA) to determine what bacteria are present at each sample site. Processed tables can be found in https://qiita.ucsd.edu/analysis/description/16572/ and https://qiita.ucsd.edu/analysis/description/15795/. Amplicon sequences were

demultiplexed and quality controlled using the defaults as provided by QIIME 1.9.1. The primary sOTU table was generated using Qiita which uses deblurring[26].

3.2.5 Data processing of MS and 16S data

Raw MS data was converted to .mzXML formatting using Bruker Data Analysis 4.1 software and uploaded into MZmine(Pluskal et al. 2010). Peak detection thresholds for MS1 and MS2 were set at 5.0E3 and 1.0E3 respectively. Chromatogram builder thresholds were set at a minimum height of 1.0E4 with an isolation tolerance of 0.01 min and m/z tolerance of 20 ppm. Chromatogram deconvolution was performed with a minimum peak height of 3.0E3 and a peak duration tolerance of 0.1-1 min. MS2 scan pairing was performed at 0.01 da and and RT scan pairing was performed at 0.4 min. Chromatograms were deisotoped with a RT tolerance of 0.2 min, an m/z tolerance of 20 ppm and selected to be have a maximum charged state of $[M+2H]^{2+}$. Peaks were aligned with up to a 20 ppm error, with 75% of the alignment priority given to m/z accuracy and 25% to 0.5 min RT overlap. Gap-filling was performed allowing a 10% intensity tolerance, a 20 ppm m/z tolerance, and a 0.4 min RT tolerance. Lastly, duplicate peak filtering was performed with a m/z tolerance of 0.01 da and a RT tolerance of 10 min. Aligned peaks were exported as file.csv and file.mgf for further analysis.

3.2.6 Data analysis of metabolome and microbiome

Multivariate analysis was performed using multiple software tools. Metabolomic principle coordinate analysis plots were constructed by importing file.csv from data

processing into Dorrestein Lab cluster application. Canberra distance methods were using with HCA grouping methods. Probabilistic quotient normalization of 30% of the samples was used. Plots were visualized in Emperor visualization[27]. PCoA plots for microbiome datasets were constructed in Qiita by using the processed un trimmed 100 rarefied sample set. Ordinates were directly observed on the Qiita interface. Alpha diversity experiments were also performed and observed in Qiita using the untrimmed waiopae BIOM table (https://qiita.ucsd.edu/analysis/description/15795/) rarerified to 100. An adonis permanova analysis was performed using vegan package in R studio[28].

Random Forest analysis of both the MZmine output data and Qiita OTU feature table was performed in Metaboanalyst[29,30] . Peak intensity tables were uploaded and normalized by sum by mean intensity values. Sum normalization was performed on each sample and mean-centered data scaling was also done. Samples were subjected to a Random Forest multivariate test.

Global Natural Products Social Molecular Networking[31] was used to identify MS/MS spectral matches of molecular features in each sample set. Parameters for creating molecular network include a parent mass tolerance=0.1 Da, Min matched Peak=6, Ion Tolerance=0.5, Score Threshold=0.80 with default advance and filter search options (http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0a0a825f41da4781a374cfb609283097). Using these parameters, 301,113 MS/MS spectra were merged into 21,732 nodes. Raw data were uploaded to http://gnps.ucsd.edu and cataloged with a massive MSV00008228. Networking parameters can be found with GNPS ID:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=940932e7fd8b42fc8f03f8b3638f0348

Molecular feature visualization was performed to visualize microbiome and metabolome features on 3d coral model. The open-source web application `ili[32] was used by incorporating feature finding tables and STereoLithography (.stl) file, OBJ 3D file (.obj), and Material Library File (.mtl) created in Photoscan.

3.3 Results

3.3.1 Coral metabolome

The metabolome of the coral community was characterized by introducing acquired LC-MS/MS data into GNPS and visualized in Cytoscape[33]. Using networking parameters described above, 64,804 spectra were considered to create nodes or unique molecular ions based on their MS2 fragmentation patterns (Figure 2). Of the 5,004 generated nodes, 205 (~4%) had matching fragmentation patterns (min match peak=4, cos score=0.70) to the GNPS molecular database (light blue nodes) and were issued level three annotations in accordance with the 2007 metabolomics standard initiative[34]. GNPS provided a FDR value of 0.01 with molecular annotations above 0.63 cos score value (Supplemental figure 2). Unannotated nodes are colored in gray.

Nodes boarders are colored based on their species of origin. A majority of the detected nodes were unannotated and shared throughout the metabolome as indicated by gray nodes with dark orange border (Figure 2). Less than 1% of the molecular network is comprised of individual nodes that are unique to species and only 1 cluster

shows specificity to a coral species. Molecular networking dimer cluster annotated by a parent mass of 364.104 m/z and 348.108 m/z were found to originate from *P. lobata*, but remain unannotated and without chemical similarity to known GNPS libraries.

The annotated metabolome of shared molecules among the sampled coral species include long and short chain fatty acids, steroids, aromatics, sugars, phospholipids, nucleosides, amino acids, chlorophyll, a pesticide and 3 unlabelled but structurally defined cyclic compounds (Figure 2). Fatty acids include hexadienoic acid, linoleic acid, pinolenic acid, docosenamide, docosatetraenoic acid, dicosatetraenoic acid, eicosanoids, monolinolein, pentanoic acid, and propanoic acid. Steroids include androsterone and campesterol. Sugar constituents include glucuronic acid and glucocerebroside. Phospholipids include phosphatidylethanolamine,phosphatidylcholine,lyso-PAF, 1-octadecyl-sn-glycero-3-phosphocholine, palmitoyl-sphingosine and phosphatidylserine. Thymidine nucleoside and decalactone were also annotated as well as a pesticide known as fuberidazole. A degradation product of chlorophyll was annotated as pheophytin A. Amino acids acetyl-L-lysine and tryptophan were annotated, while the four structurally defined but unlabelled clusters have m/z values of 628.295, 415.212, and 247.167.

3.3.2 Coral microbiome

Closed-reference OTU sequences were used to characterize the microbiome taxonomy of the sampled coral assemblage to species level classification. Relative

frequency of 16S taxa were visualized against all samples (Figure 3). The top 10 most abundant genus present across the coral community in descending order include *Rhodobacterales* which is found in all of the samples including the water column. *Oceanospirillales* appears to dominate *P. lobata* tissue samples which is consistent across many *Porites* coral communities worldwide[35]. *Rhizobiales* appear to be distributed throughout all samples except for *P. lobata coral*. *Chlorophyta* was present in *M. capitata* and *P. lobata* and wasabsent in *M. flabellata* samples. *Stramenopiles* is seen to be consistently present across all samples but absent in a subset of diseased GA tissue samples from *M. capitata*. *Oscillatoriales, Pirellulales, Acidimicrobiales*, and *Verrucomicrobiales* were present across all species except for *P. lobata*. A Complete list of annotated 16S OTU sequences up to species level classification can be found in the supplemental information.

3.3.3 molecular diversity across coral assemblages

Metabolomic diversity (H) was significantly lower (pvalue) in *P. lobata* compared to both species of *Montipora.* Diversity (H) was also lower in GA-affected *M. capitata* tissue compared to healthy tissues of *M. capitata* and *M. flabellata*. Significant differences (pvalue<0.0001) in the microbial diversity (H) was found when comparing diversity (H) among all sampled corals (Figure 4). Kruskal-Wallis pairwise comparisons among the species revealed that GA-affected *M. capitata* tissue had significantly less diversity when compared to healthy *M. flabellata* and *M. capitata* tissue (pvalue<0.05).

Similarly, *P. lobata* tissue had significantly less diversity than *M. flabellata* and *M. capitata* tissue (pvalue<0.0001).

Adonis (PERMANOVA) test on LC-MS/MS and 16S OTUs feature tables identified significant beta-diversity among species for both the metabolome (PERMANOVA<0.001) and the microbiome (PERMANOVA<0.001). PCoA bray curtis ordination plots were created to visualize the differences identified in PERMANOVA analysis (Figure 5). Groups were divided by *M. capitata*, *M. flabellata*, and *P. lobata coral species*.

Random forest classification method identified features that describe the metabolome and microbiome species variation with an out-of-bag (OOB) error of 0.0056 and 0.362 respectively (Supplemental figure 3). Despite random forest providing an inventory of up to species level microbial features that describe microbiome variation, GNPS molecular networking library annotations was only able to identify 1-octadecyl-sn-glycero-3-phosphocholine of the top 15 features to describe metabolomic variance among the sample coral species. Additionally, random forest was able to identify features that describe variation between GA-affected *M. capitata* tissue and healthy *M. capitata* tissue with an OOB error of 0.149 for microbiome datasets and 0.350 for metabolomic datasets (Supplemental figure 4). The top microbial taxa feature that appears to be absent in disease tissue and present in healthy tissue is classified as a cyanobacteria belonging to the family *Stramenopiles*. The only GNPS level 3 annotated molecular feature in the metabolome is pheophytin A also known to be a degradation product of chlorophyll.

3.3.4 Molecular features that describe coral growth anomaly presences and absence

The top features identified from random forest analysis that describes the difference in molecular profiles between GA-affected tissue and healthy *M. capitata* tissue was also visualized in the ili 3D cartography platform (Figure 6a-d). Pheophytin A abundances were visualized and identified. *Cyanobacteria* belonging to the family *Stramenopiles* were on the 3D coral assemblage at each sample location to visualize the higher abundance at all sample locations compared to the sampled GA-affected tissue (Figure 6a,b). Similarly, pheophytin A abundances were undetectable on GA-affected sample site locations when visualized with the ili platform (Figure 6c,d). A fully interactive visualization can be viewed on the ili platform interface for the metabolome (https://ili.embl.de/?ftp://massive.ucsd.edu/MSV000080242/updates/2018-04-24_ckapono_bf2e7d5e/other/Waiopae_biome_ili.obj;ftp://massive.ucsd.edu/MSV000080242/updates/2018-04-24_ckapono_d1c36f12/other/Annotated_GNPS_featurew_for_ILI_test_OBJ_MARCH218.csv;ftp://massive.ucsd.edu/MSV000080242/updates/2018-04-24_ckapono_2e7bcdb8/other/coral_metabolome.json) and the microbiome (https://ili.embl.de/?ftp://massive.ucsd.edu/MSV000080242/updates/2018-04-24_ckapono_bf2e7d5e/other/Waiopae_biome_ili.obj;ftp://massive.ucsd.edu/MSV000080242/updates/2018-04-24_ckapono_bf2e7d5e/other/ILI_100rare_OTU_taxa_level-7.csv;ftp://massive.ucsd.edu/MSV000080242/updates/2018-04-24_ckapono_2e7bcdb8/other/coral_microbiome.json).

3.4 Discussion

Phenotypic classification, genetic sequencing techniques, and visual surveys have been used as conventional methods for biological characterizations of coral reefs. 3D molecular profiling offers a complementary investigatory technique that can collate these approaches to better characterize differences across species and health states of coral assemblages. 3D profiling of the metabolomic and microbial diversity of a coral assemblage provided useful information about the molecular diversity of several coral species at Waiōpae. We identified differences in the composition and diversity of microbial and metabolomic profiles among *M. captiata*, *M. flabelatta*, and *P. lobata* corals. Applying a multivariate statistical approach also provided evidence of microbial and metabolomic differences between healthy and diseased tissues.. Integrating genomic and metabolomic data into a 3D platform provides a powerful tool for spatial analyzing molecular profiles across the dynamic landscape of coral colonies.

Utilizing a multivariate statistical enabled identification of key molecular features that were significantly different among the sampled coral species and GA-affected tissue. Random forest (RF) classification found specific bacteria and molecules associated with each species and GA-affected tissue. Although these features identified by random forest individually represent less than 1% of the total classification accuracy, the top 15 features from the classification describe 4% of the microbiome variance and 10% of the metabolome variance. A slightly higher level of variance was described when specifically comparing GA-affected and healthy tissue from *M. capitata* coral.

Individual microbial features and features detected with RF classification describe 2% of the microbiome variance and 5% of metabolome variance all coral samples. These findings suggests that specific groups of molecules and microbes may need to be targeted to characterize profiles that are unique to individual coral species and are health-states.

The process of unraveling coral metabolomes is in its infancy, but as profiling databases continue to advance, an ever-growing portion of the cryptic metabolome is being unearthed. Limitations of molecular annotation in LC-MS/MS metabolomic studies often lead to thousands of unannotated spectra. GNPS molecular networking has the capability to overcome these issues by implementing MS/MS spectral matching which provides level 3 annotation as described by the 2007 metabolomic initiative[34]. As the GNPS database continues to grow beyond its 10,000 molecular annotations, the potential annotation resolution will improve dramatically. Due to the mechanism of annotation through MS2 spectral matching, false annotations are possible. False discovery rates (FDR) have been implemented into GNPS to increase confidence and each level 3 metabolomic annotation may require molecular isolation or comparison to standards (Supplemental Figure 2). As the GNPS network continues to grow and detection accuracy improves it will become increasingly useful to utilize a metabolomic approach for characterizing the biology of marine organisms.

Identifying specific compounds within the coral tissue provides useful insight for understanding the mechanics of diseased coral tissue. Sterols such as campesterol have been previously found in tissue tumors[36] and appear to be to be in high abundance

in one of the locations of GA-affected tissue when analyzed in a 3D framework (Supplemental figure 5). Saccharides are a major composition of coral inter skeletal physiology[37] and have shown to be present across all of the coral community in high abundances . Sugar constituents include glucuronic acid and glucocerebroside. Phospholipids include phosphatidylethanolamine, phosphatidylcholine, lyso-PAF, 1-octadecyl-sn-glycero-3-phosphocholine, palmitoyl-sphingosine and phosphatidylserine. Comparing the abundance and diversity of these compounds among the samples provides a useful platform for identifying molecules known to be associated with other similar disease states. Considering the biology and ecology of coral diseases are still poorly understood[38], it is valuable to be able to search for molecules known to be associated with other animal diseases to improve our characterization of conditions affecting corals.

Fatty acid and phospholipid molecules detected within the metabolome (Figure 2) of the Waiōpae coral assemblage are consistent with previously identified primary metabolites in sceleterian corals, soft coral, sponges and dinoflagellates. These molecules have also been linked to elicit broad host immune responses[39,40] and appear to be the main component to the polar metabolome of marine metazoans. Identifying the widespread spatial distribution (Figure 5) of this interesting class of molecules may help investigators prioritize the need to further understand their ecological role.

No significant molecular differences were detected among the healthy coral tissue of *Montipora captiata*, yet tissue affected by the GA disease exhibited significantly decreased molecular diversity (Figure 2). While more research and an

increased sample size will help to validate this relationship, the decrease in molecule abundance within GA-affected tissue suggests this disease compromises metabolic activity in the organism or promotes a disproportional production of specialized metabolites. Microbial diversity was found to be significantly lower in *P. lobata* than the other sampled coral species (Figure 2 and 3) and supports previous research that also observed lower bacterial diversity across *Porites* coral species in other locations[4]. The lower microbial diversity found in the GA-affected tissue is also consistent with many microbial studies investigating the health of organisms in an array of environmental and ecological settings[41–43]. The significantly decreased diversity in both microbial and molecular features in GA-affected and *P. lobata* tissue highlights the need for more research using these techniques to determine if molecular profiles can serve as biomarkers for discriminating among species and health-states of coral.

Pheophytin A was one of the top 15 features that was successfully identified through the GNPS spectral matching algorithm. Its spatial distribution was widespread throughout the sampled coral assemblage but was absent from *M. capitata* GA-affected tissue (Figure 6). As pheophytin is a degradation product of chlorophyll, this finding illustrates the lack of photosynthesis occurring in GA-affected tissue compared to healthy tissue and provides more evidence of consequences this disease has on the organism's overall metabolism. Light transforming symbiotic microorganisms within coral tissue provide much needed nutrients for the host[44,45]. It is unclear to whether the lack of photosynthesis contributes to or is a result of diseased tissue, but supports prior research indicating distinct photophysiological reduction in coral tissue affected by this

disease[17]. This finding provides another useful example of how metabolomic biomarkers can identify physiological impairment associated with diseased coral tissue.

Algal DNA was detected on several coral samples. This is likely due to the similar genetic homology of the V4 region of the 16S ribosome to chloroplast.. Although some have viewed chloroplast contamination as a problem[46], detecting increased amounts of microboring algae[47] within GA-affected tissue is an interesting finding. Previous studies have investigated the relationship of turf algae growth on GA-tissue at Waiōpae and found growth to occur on both diseased and healthy tissue[47,48]. The detection of *Ulvophyceae* suggests this algae has found an opportunity to infiltrate colony defense and begin algal turf succession on live coral tissue. Surprisingly, the higher abundance of microboring *Ulvophyceae* is inversely proportional to the amount of chloroplast detected in the samples. This may be associated with a lack of polyps and mesenterial filaments in GA-affected tissue[49], which could potentially reduce the ability sweep away algae attempting to settle and grow on live coral tissue. Further investigation on the molecular profiles of this disease may help to shed light on the biology and long-term impacts of GA disease to the affected coral host.

Other bacteria from the genus have also been identified as key descriptors of the variance between health and disease tissues. Although Rodobaterecae has been linked to coral disease[50,51], it appears that the absence of this genus of bacteria is linked to diseased tissue. More investigations need to be performed to investigate coral microbiome communities surrounding GA disease tissue and perhaps biases with chloroplast contamination may need to be taken into account.

As in any study, inherent biases within sampling, extraction and processing procedures may influence the observed values of shannon diversity. Although microbiome analytics are helping to overcome such issues, investigators are constantly searching for improved metabolomic extraction methods that will provide a more comprehensive and accurate chemical profile. One recent approach adopted in marine chemical profiling is the integration of multiple metabolomic data acquisition methods such as GC-MS, LC-DAD-IT, LC-DAD-TOF and MALDI-TOF/TOF into a single profiling study[8]. Future studies should consider comparing multiple acquisition techniques to assess variability in molecular profiles that may be driven by methodological differences.

As reef communities around the world suffer declines due to anthropogenic disturbances and disease, new technologies such as 3D molecular profiling offer insightful and complementary investigative potential. Collating traditional coral profiling tools into a synthesized analysis has the capability and scalability to improve our understanding of the biological underpinnings that drive the ecology of these valuable coral reef ecosystems. As demonstrated in this study, 3D molecular profiling can be a powerful and reproducible, and tool that can enhance future research of natural systems.

3.5 Appendix of figures for chapter 3

Figure 1. Coral samples were collected from and overlapping images of the coral assemblage were taken from oblique and planar angle (a). Agisoft Photoscan was used to create a 3D model of the coral assemblage and sampling coordinates were annotated using Geomagic Wrap software. (Samples were subjected to both metabolite and DNA extractions (b). Samples were subjected to LC-MS/MS and 16 rRNA V4 amplicon sequencing (c). Raw mass spectra and sequencing outputs were converted and assembled to mzXML and FASTQ formats respectively. GNPS, Metaboanalyst, MZmine 2,Qiita and QIIME software was used to process data (d). Data analysis was performed in visualization Ili, EMPeror, Cytoscape, Qiita and Metabolanalyst graphic user interfaces (e).

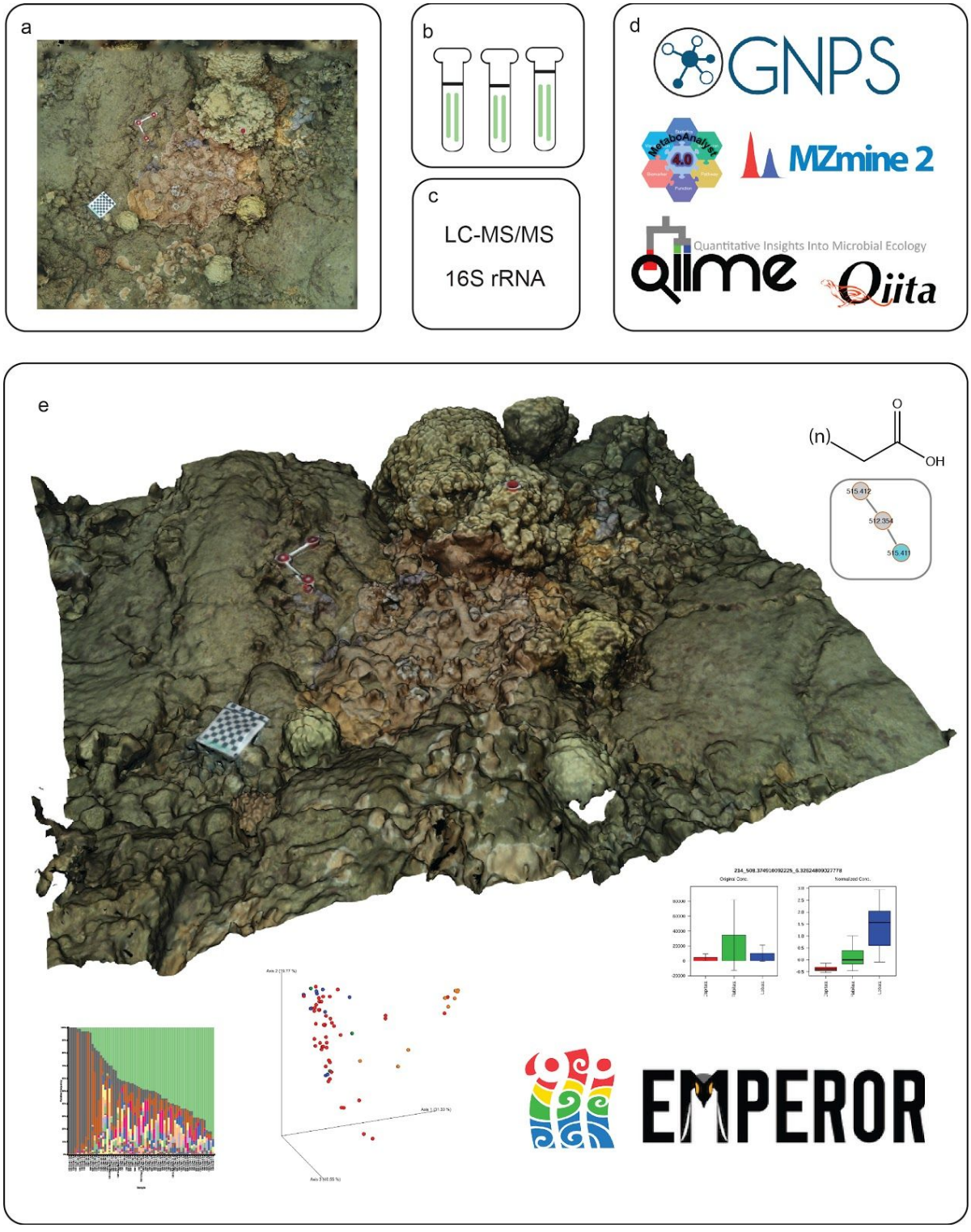

a
b
c
LC-MS/MS
16S rRNA
d
GNPS
MetaboAnalyst 4.0
MZmine 2
Quantitative Insights Into Microbial Ecology
qiime
Qiita
e
(n)
O
OH
515.412
512.354
515.411
EMPEROR

Figure 2. Level three annotations are highlighted by blue nodes. Unannotated molecules are represented by gray nodes. Light blue node border represents molecules only found in M. flabellata. Dark blue node border represents molecules found in all species except M. flabellata. Light orange nodes represents molecules only found in M. capitata. Dark orange nodes represent molecules found in all species except P. lobata. Light green borders are molecules found only in P. lobata. Dark green borders found in all species except M. flabellata.

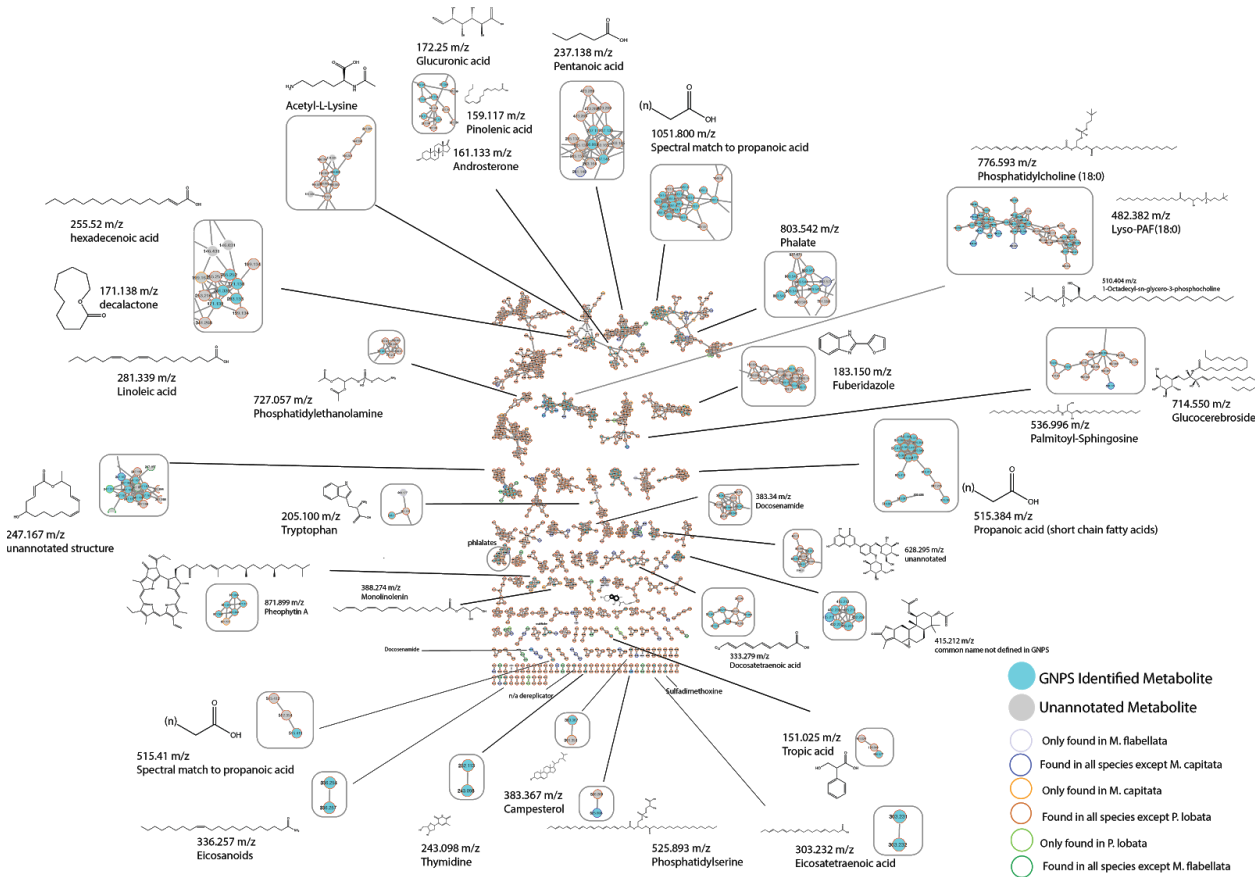

172.25 m/z
Glucuronic acid
237.138 m/z
Pentanoic acid
Acetyl-L-Lysine
159.117 m/z
Pinolenic acid
161.133 m/z
Androsterone
1051.800 m/z
Spectral match to propanoic acid
776.593 m/z
Phosphatidylcholine (18:0)
482.382 m/z
Lyso-PAF(18:0)
510.404 m/z
1-Octadecyl-sn-glycero-3-phosphocholine
255.52 m/z
hexadecenoic acid
803.542 m/z
Phalate
171.138 m/z
decalactone
281.339 m/z
Linoleic acid
727.057 m/z
Phosphatidylethanolamine
183.150 m/z
Fuberidazole
536.996 m/z
Palmitoyl-Sphingosine
714.550 m/z
Glucocerebroside
247.167 m/z
unannotated structure
205.100 m/z
Tryptophan
383.34 m/z
Docosenamide
515.384 m/z
Propanoic acid (short chain fatty acids)
phalates
628.295 m/z
unannotated
871.899 m/z
Pheophytin A
388.274 m/z
Monolinolenin
415.212 m/z
common name not defined in GNPS
333.279 m/z
Docosatetraenoic acid
Docosenamide
n/a dereplicator
Sulfadimethoxine
515.41 m/z
Spectral match to propanoic acid
151.025 m/z
Tropic acid
383.367 m/z
Campesterol
336.257 m/z
Eicosanoids
243.098 m/z
Thymidine
525.893 m/z
Phosphatidylserine
303.232 m/z
Eicosatetraenoic acid
GNPS Identified Metabolite
Unannotated Metabolite
Only found in M. flabellata
Found in all species except M. capitata
Only found in M. capitata
Found in all species except P. lobata
Only found in P. lobata
Found in all species except M. flabellata

Figure 3. 16S rRNA v4 amplicon sequences were taken from each sample and organized by genus-level annotated closed reference OTU taxonomy.

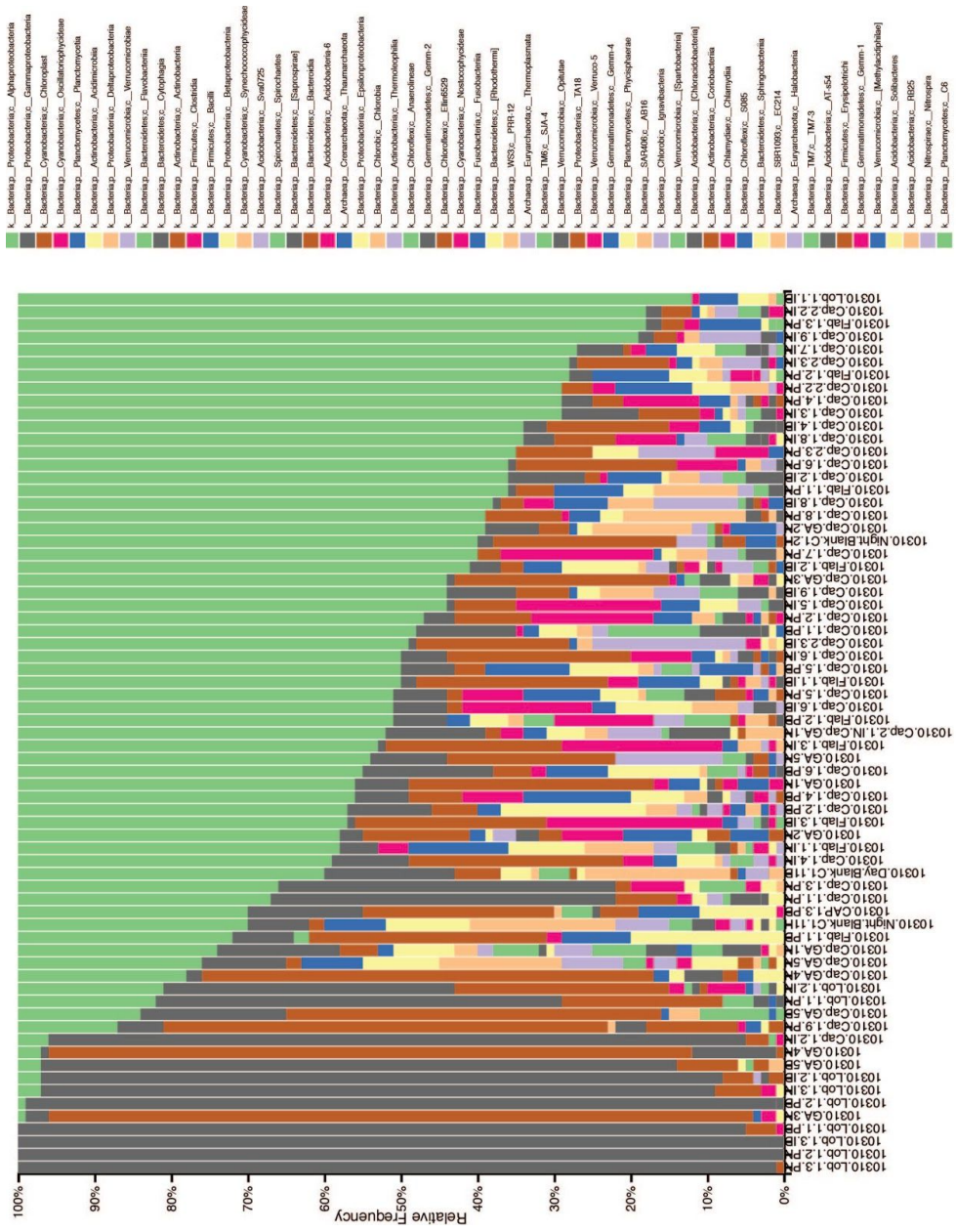
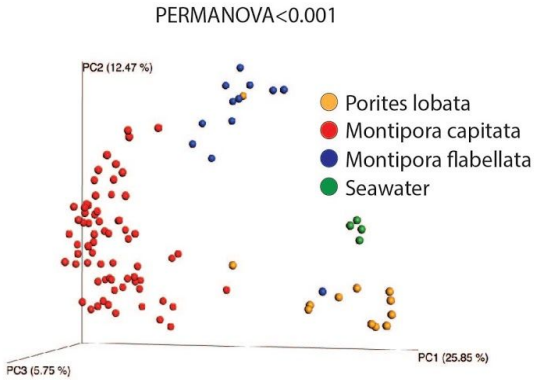

K__Bacteria;p__Proteobacteria;c__Alphaproteobacteria
K__Bacteria;p__Proteobacteria;c__Gammaproteobacteria
K__Bacteria;p__Cyanobacteria;c__Chloroplast
K__Bacteria;p__Cyanobacteria;c__Oscillatoriophycideae
K__Bacteria;p__Planctomycetes;c__Planctomycetia
K__Bacteria;p__Actinobacteria;c__Acidimicrobiia
K__Bacteria;p__Proteobacteria;c__Deltaproteobacteria
K__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae
K__Bacteria;p__Bacteroidetes;c__Flavobacteriia
K__Bacteria;p__Actinobacteria;c__Cytophagia
K__Bacteria;p__Actinobacteria;c__Actinobacteria
K__Bacteria;p__Firmicutes;c__Clostridia
K__Bacteria;p__Firmicutes;c__Bacilli
K__Bacteria;p__Proteobacteria;c__Betaproteobacteria
K__Bacteria;p__Cyanobacteria;c__Synechococcophycideae
K__Bacteria;p__Acidobacteria;c__Sva0725
K__Bacteria;p__Spirochaetes;c__Spirochaetes
K__Bacteria;p__Bacteroidetes;c__[Saprospirae]
K__Bacteria;p__Bacteroidetes;c__Bacteroidia
K__Bacteria;p__Acidobacteria;c__Acidobacteria-6
K__Archaea;p__Crenarchaeota;c__Thaumarchaeota
K__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria
K__Bacteria;p__Chloroflexi;c__Chloroflexi
K__Bacteria;p__Actinobacteria;c__Thermoleophilia
K__Bacteria;p__Chloroflexi;c__Anaerolineae
K__Bacteria;p__Gemmatimonadetes;c__Gemm-2
K__Bacteria;p__Chloroflexi;c__Ellin6529
K__Bacteria;p__Cyanobacteria;c__Nostocophycideae
K__Bacteria;p__Fusobacteria;c__Fusobacteriia
K__Bacteria;p__Bacteroidetes;c__[Rhodothermi]
K__Bacteria;p__WS3;c__PRR-12
K__Archaea;p__Euryarchaeota;c__Thermoplasmata
K__Bacteria;p__TM6;c__SJA-4
K__Bacteria;p__Verrucomicrobia;c__Opitutae
K__Bacteria;p__Proteobacteria;c__TA18
K__Bacteria;p__Verrucomicrobia;c__Verruco-5
K__Bacteria;p__Gemmatimonadetes;c__Gemm-4
K__Bacteria;p__Planctomycetes;c__Physcisphaerae
K__Bacteria;p__SAR406;c__AB16
K__Bacteria;p__Chloroflexi;c__Ignavibacteriae
K__Bacteria;p__Verrucomicrobia;c__[Spartobacteria]
K__Bacteria;p__Acidobacteria;c__[Chloracidobacteria]
K__Bacteria;p__Actinobacteria;c__Coriobacteriia
K__Bacteria;p__Chlamydiae;c__Chlamydiae
K__Bacteria;p__Chloroflexi;c__S085
K__Bacteria;p__Bacteroidetes;c__Sphingobacteriia
K__Bacteria;p__SBR1093;c__EC214
K__Archaea;p__Euryarchaeota;c__Halobacteria
K__Bacteria;p__TM7;c__TM7-3
K__Bacteria;p__Acidobacteria;c__AT-S54
K__Bacteria;p__Firmicutes;c__Erysipelotrichi
K__Bacteria;p__Gemmatimonadetes;c__Gemm-1
K__Bacteria;p__Verrucomicrobia;c__[Methylacidiphilae]
K__Bacteria;p__Acidobacteria;c__Solibacteres
K__Bacteria;p__Acidobacteria;c__RB25
K__Bacteria;p__Nitrospirae;c__Nitrospira
K__Bacteria;p__Planctomycetes;c__OM
Relative Frequency
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Sample
10310.Lob.1.3.PB
10310.Lob.1.2.PB
10310.Lob.1.3.IB
10310.Lob.1.1.PB
10310.GA.3N
10310.Lob.1.2.PB
10310.Lob.1.3.IN
10310.Lob.1.2.IB
10310.GA.5B
10310.GA.4N
10310.Cap.1.2.IN
10310.Cap.1.9.PB
10310.Cap.GA.5B
10310.Lob.1.1.PB
10310.Lob.1.2.IN
10310.Cap.GA.4N
10310.Cap.GA.5N
10310.Cap.GA.1N
10310.Fab.1.1.PB
10310.CAP.1.3.PB
10310.Night.Blank.C1.11N
10310.Cap.1.1.PB
10310.Cap.1.3.PB
10310.Day.Blank.C1.11B
10310.Fab.1.1.IN
10310.Cap.1.4.IN
10310.GA.2N
10310.Fab.1.3.IB
10310.Cap.1.2.PB
10310.Cap.1.4.PB
10310.GA.1N
10310.GA.5N
10310.Cap.1.6.PB
10310.Fab.1.3.IN
10310.Cap.2.1.IN.Cap.GA.1N
10310.Fab.1.2.PB
10310.Cap.1.6.IB
10310.Cap.1.5.PN
10310.Fab.1.1.IB
10310.Cap.1.5.PB
10310.Cap.1.6.IN
10310.Cap.2.3.IB
10310.Cap.1.1.PB
10310.Cap.1.2.PN
10310.Cap.1.5.IN
10310.Cap.1.9.IB
10310.Fab.1.2.IB
10310.Cap.GA.3N
10310.Cap.1.7.PN
10310.Night.Blank.C1.2N
10310.Cap.GA.2N
10310.Cap.1.8.PN
10310.Fab.1.1.PN
10310.Cap.1.2.IB
10310.Cap.1.6.PN
10310.Cap.2.3.PN
10310.Cap.1.4.IB
10310.Cap.1.3.IN
10310.Cap.1.4.PN
10310.Fab.1.2.PN
10310.Cap.2.3.IN
10310.Cap.1.7.IN
10310.Cap.1.9.IN
10310.Fab.1.3.PN
10310.Cap.2.2.IN
10310.Day.Blank.C1.1B

Figure 4. Kruskal-Wallis test was performed to compare shannon index of the metabolome (a) and microbiome (b) of blue rice, brown rice, and lobe species of coral. Brown rice coral with GA lesions were also compared.

a
shannon
p-value < 1.0e-7
Montipora flabelatta (n=12)
Montipora capitata (n=56)
Montipora capitata GA (n=10)
Porites lobata (n=11)
Seawater (n=4)
Coral
b
shannon
p-value <1.0e-4
Montipora flabelatta (n=9)
Montipora capitata (n=41)
Montipora capitata GA (n=6)
Porites lobata (n=10)
Sea water (n=3)
Coral

Figure 5. PCoA EMPeror visualization of the bray distance method comparing the metabolome (a) and microbiome (b) of *M. capitata*, *M. flabelatta*, and *P. lobata*.

a


PERMANOVA<0.001
PC2 (12.47 %)
PC3 (5.75 %)
PC1 (25.85 %)
Porites lobata
Montipora capitata
Montipora flabellata
Seawater

b

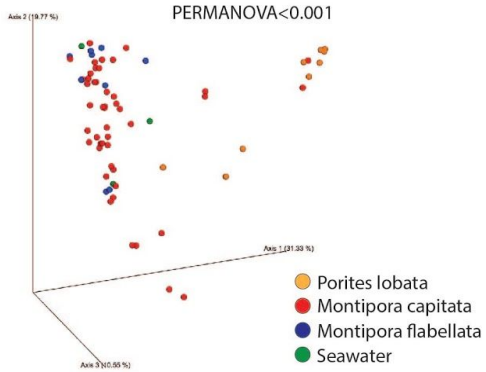
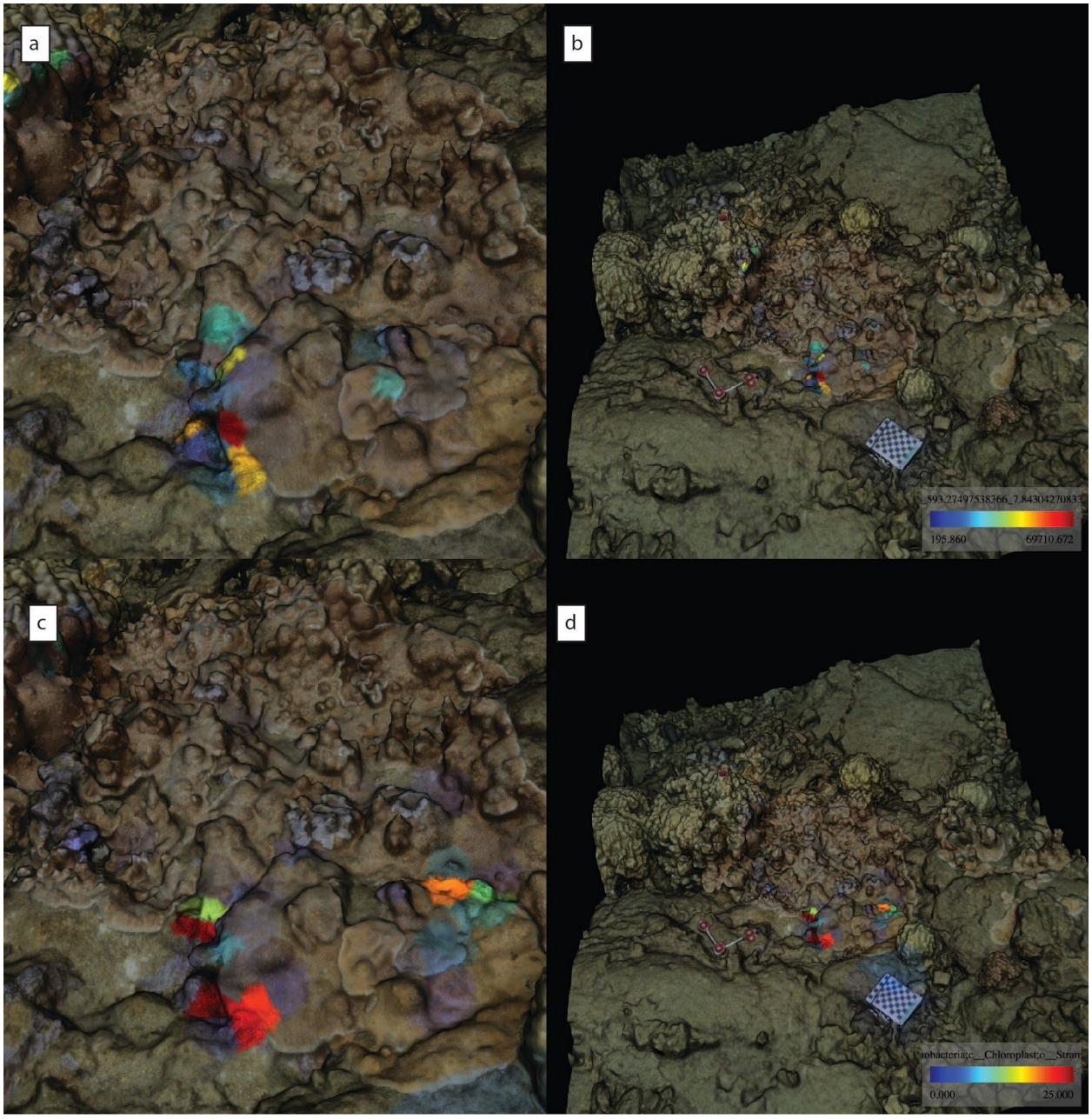

PERMANOVA<0.001
Porites lobata
Montipora capitata
Montipora flabellata
Seawater

Figure 6. 3D visualization of key features that drive variance between healthy and diseased M. capitata tissue. Ili visualization of the *Stramenopiles* distribution at across coral assemblage (a,b). Ili visualization of the pheophytin A distribution across the coral assemblage (c,d).


a
b
195.860
69710.672
c
d
0.000
25.000

3.6 Acknowledgments

The work presented in Chapter 3 is the result of collaborations with the Takabayashi lab at the University of Hawaii at Hilo as well as the Knight lab at the University of California San Diego. Misaki Takabayashi and John Burns both help to provide preliminary data and insight on the growth anomaly condition in Kapoho, Hawaii. Takabayashi lab member John Burns provided assistance in the photo collection and the creation of the coral assemblage 3D model. Takabayashi lab members, Lauren Kapono and Kanoe Steward help with sample collection. Greg Humphrey from the Knight lab performed DNA extractions and prepared them for Illumina sequencing. Gail Ackerman provided guidance during upload and processing of coral metadata in Qiita. Chapter 3 is currently being prepared for submission of the material. Kapono, C.A., Burns, J.H.R., Ackermann, G., Humphrey, G., Takabayashi, M., Knight, R., Dorrestein, P.C. The dissertation author was the primary investigator and author of this manuscript.

3.7 References

1. Lamb JB, Willis BL, Fiorenza EA, Couch CS, Howard R, Rader DN. Plastic waste associated with disease on coral reefs. Science. 2018 Jan 26;359(6374):460–2.
2. Roche RC, Williams GJ, Turner JR. Towards Developing a Mechanistic Understanding of Coral Reef Resilience to Thermal Stress Across Multiple Scales. Current Climate Change Reports. 2018;4(1):51–64.
3. F. R, Rohwer F, Breitbart M, Jara J, Azam F, Knowlton N. Diversity of bacteria associated with the Caribbean coral Montastraea franksi. Coral Reefs. 2001;20(1):85–91.
4. Rohwer F, Seguritan V, Azam F, Knowlton N. Diversity and distribution of coral-associated bacteria. Mar Ecol Prog Ser. 2002;243:1–10.

5. Bourne DG, Garren M, Work TM, Rosenberg E, Smith GW, Harvell CD. Microbial disease and the coral holobiont. Trends Microbiol. 2009 Dec;17(12):554–62.

6. McDevitt-Irwin JM, Baum JK, Garren M, Vega Thurber RL. Responses of Coral-Associated Bacterial Communities to Local and Global Stressors. Frontiers in Marine Science [Internet]. 2017;4. Available from: http://dx.doi.org/10.3389/fmars.2017.00262

7. Gordon BR, Leggat W. Symbiodinium—Invertebrate Symbioses and the Role of Metabolomics. Mar Drugs. 2010;8(10):2546–68.

8. Costa-Lotufo LV, Carnevale-Neto F, Trindade-Silva AE, Silva RR, Silva GGZ, Wilke DV. Chemical profiling of two congeneric sea mat corals along the Brazilian coast: adaptive and functional patterns. Chem Commun . 2018 Feb 20;54(16):1952–5.

9. Sogin EM, Putnam HM, Anderson PE, Gates RD. Metabolomic signatures of increases in temperature and ocean acidification from the reef-building coral, Pocillopora damicornis. Metabolomics [Internet]. 2016;12(4). Available from: http://dx.doi.org/10.1007/s11306-016-0987-8

10. Hartmann AC, Petras D, Quinn RA, Protsyuk I, Archer FI, Ransome E. Meta-mass shift chemical profiling of metabolomes from coral reefs. Proc Natl Acad Sci U S A. 2017 Oct 31;114(44):11685–90.

11. J. B, Bythell J, Pan P, Lee J. Three-dimensional morphometric measurements of reef corals using underwater photogrammetry techniques. Coral Reefs. 2001;20(3):193–9.

12. Leon JX, Roelfsema CM, Saunders MI, Phinn SR. Measuring coral reef terrain roughness using “Structure-from-Motion” close-range photogrammetry. Geomorphology . 2015;242:21–8.

13. Burns J, Delparte D, Gates RD, Takabayashi M. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. PeerJ. 2015 Jul 7;3:e1077.

14. Storlazzi CD, Dartnell P, Hatcher GA, Gibbs AE. End of the chain? Rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (SfM) technology. Coral Reefs. 2016;35(3):889–94.

15. Burns JHR, Delparte D, Kapono L, Belt M, Gates RD, Takabayashi M. Assessing the impact of acute disturbances on the structure and composition of a coral community using innovative 3D reconstruction techniques. Methods in Oceanography. 2016;15-16:49–59.

16. Burns JHR, Rozet NK, Takabayashi M. Morphology, severity, and distribution of growth anomalies in the coral, Montipora capitata, at Waiōpae, Hawaii. Coral Reefs.

2011;30(3):819–26.

17. Burns JHR, Gregg TM, Takabayashi M. Does coral disease affect symbiodinium? Investigating the impacts of growth anomaly on symbiont photophysiology. PLoS One. 2013 Aug 14;8(8):e72466.

18. Spies NP, Takabayashi M. Expression of galaxin and oncogene homologs in growth anomaly in the coral Montipora capitata. Dis Aquat Organ. 2013 Jun 13;104(3):249–56.

19. Burns JHR, Alexandrov T, Ovchinnikova K, Gates RD, Takabayashi M. Data for spatial analysis of growth anomaly lesions on coral colonies using 3D reconstruction techniques. Data Brief. 2016 Dec;9:460–2.

20. Bruno F, Gallo A, De Filippo F, Muzzupappa M, Petriaggi BD, Caputo P. 3D documentation and monitoring of the experimental cleaning operations in the underwater archaeological site of Baia (Italy). In: 2013 Digital Heritage International Congress (DigitalHeritage) [Internet]. 2013. Available from: http://dx.doi.org/10.1109/digitalheritage.2013.6743719

21. McCarthy J, Benjamin J. Multi-image Photogrammetry for Underwater Archaeological Site Recording: An Accessible, Diver-Based Approach. J Marit Archaeol. 2014;9(1):95–114.

22. Kersten TP, Lindstaedt M. Automatic 3D Object Reconstruction from Multiple Images for Architectural, Cultural Heritage and Archaeological Applications Using Open-Source Software and Web Services Automatische 3D-Objektrekonstruktion aus digitalen Bilddaten für Anwendungen in Architektur, Denkmalpflege und Archäologie durch open-source Software und Webservices. Photogrammetrie - Fernerkundung - Geoinformation. 2012;2012(6):727–40.

23. Gordon BR, Leggat W, Motti CA. Extraction protocol for nontargeted NMR and LC-MS metabolomics-based analysis of hard coral and their algal symbionts. Methods Mol Biol. 2013;1055:129–47.

24. Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J. The Earth Microbiome Project: Meeting report of the “1st EMP meeting on sample selection and acquisition” at Argonne National Laboratory October 6th 2010. Stand Genomic Sci. 2010;3(3):249–53.

25. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012 Aug;6(8):1621–4.

26. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems [Internet]. 2017 Mar;2(2). Available from:

http://dx.doi.org/10.1128/mSystems.00191-16

27. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11(1):395.

28. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience. 2013 Nov 26;2(1):16.

29. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14(6):927–30.

30. Xia J, Wishart DS. Metabolomic Data Processing, Analysis, and Interpretation Using MetaboAnalyst. Curr Protoc Bioinformatics. 2011;34(1):14.10.1–14.10.48.

31. Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0--making metabolomics more meaningful. Nucleic Acids Res. 2015 Jul 1;43(W1):W251–7.

32. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016 Aug 9;34(8):828–37.

33. Protsyuk I, Melnik AV, Nothias L-F, Rappez L, Phapale P, Aksenov AA. 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. Nat Protoc. 2018 Jan;13(1):134–54.

34. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498–504.

35. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007 Sep;3(3):211–21.

36. Mark D. Speck SPD. Widespread Oceanospirillaceae Bacteria in Porites spp. J Mar Biol [Internet]. 2012;12. Available from: http://dx.doi.org/10.1155/2012/746720

37. Yamashiro H, Oku H, Onaga K, Wasaki H, Takarab K. Coral tumors store reduced level of lipids. J Exp Mar Bio Ecol. 2001;265(15):171–9.

38. Adamiano A, Goffredo S, Dubinsky Z, Levy O, Fermani S, Fabbri D. Analytical pyrolysis-based study on intra-skeletal organic matrices from Mediterranean corals. Anal Bioanal Chem. 2014 Sep;406(24):6021–33.

39. Richardson LL. Coral diseases: what is really known? Trends Ecol Evol. 1998 Nov 1;13(11):438–43.

40. Garg N, Kapono C, Lim YW, Koyama N, Vermeij MJA, Conrad D. Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures. Int J Mass Spectrom. 2015 Feb 1;377:719–717.

41. Quinn RA, Vermeij MJA, Hartmann AC, Galtier d'Auriac I, Benler S, Haas A. Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. Proc Biol Sci [Internet]. 2016 Apr 27;283(1829). Available from: http://dx.doi.org/10.1098/rspb.2016.0469

42. Clay K, Holah J. Fungal endophyte symbiosis and plant diversity in successional fields. Science. 1999 Sep 10;285(5434):1742–5.

43. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM. Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. J Infect Dis. 2008 Feb 1;197(3):435–8.

44. Keesing F, Belden LK, Daszak P, Dobson A, Harvell CD, Holt RD. Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature. 2010 Dec 2;468(7324):647–52.

45. Muscatine L, Porter JW. Reef Corals: Mutualistic Symbioses Adapted to Nutrient-Poor Environments. Bioscience. 1977;27(7):454–60.

46. Hatcher BG. Coral reef primary productivity: A beggar's banquet. Trends Ecol Evol. 1988;3(5):106–11.

47. Hanshew AS, Mason CJ, Raffa KF, Currie CR. Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. J Microbiol Methods. 2013 Nov;95(2):149–55.

48. Massé A, Domart-Coulon I, Golubic S, Duché D, Tribollet A. Early skeletal colonization of the coral holobiont by the microboring Ulvophyceae Ostreobium sp. Sci Rep. 2018 Feb 2;8(1):2293.

49. Claar DC, Takabayashi M. The effects of growth anomaly on susceptibility of Montipora capitata to turf algal overgrowth. Mar Freshwater Res. 2016;67(5):666.

50. Burns JHR, Takabayashi M. Histopathology of growth anomaly affecting the coral, Montipora capitata: implications on biological functions and population viability. PLoS One. 2011 Dec 19;6(12):e28854.

51. Roder C, Arif C, Bayer T, Aranda M, Daniels C, Shibl A. Bacterial profiling of White Plague Disease in a comparative coral species framework. ISME J. 2013;8(1):31–9.

52. Sunagawa S, DeSantis TZ, Piceno YM, Brodie EL, DeSalvo MK, Voolstra CR. Bacterial diversity and White Plague Disease-associated community changes in the Caribbean coral Montastraea faveolata. ISME J. 2009 May;3(5):512–21.

4. Chapter 4

4.1 Introduction

Over the last decade, several studies have alluded to the notion of an organism's' habitat driving molecular uniqueness[1–6], but more studies are needed to explore the role geography and behavior have on these populations[7]. The surfer biome project (SBP) looks to not only explore human environmental interaction, but do so while incorporating both traditional knowledge and "western science."

Indigenous knowledge has long asserted that humans are not separate from their surrounding environment[8–12]. A perspective that is often neglected in modern science[13–15].The Kumulipo chant(Liliuokalani 1897), a Hawaiian phylogenetic description of evolution, exemplifies traditional ideologies about the innate connection humans have with their environment. Passed down through oral history, it describes the first Hawaiian ancestors as progeny of the land. Seventeenth century Hawaiian literature also depicts genealogical relationships of human and nature. Written by Native Hawaiian scholar and environmental activist Joseph Iosepa Kahooluhi Nawahiokalaniopuu, the idea of the land being an ancestral grandmother is well documented(Nawahi 1895; Silva 2004). Furthermore, the deep connection of Hawaiians and nature was additionally illustrated in 20th century legislation. US public law 103-150 103d joint resolution 19 states that "Whereas, the health and wellbeing of the Native Hawaiian people is intrinsically tied to their deep feelings and attachment to the land."

Hee nalu, affectionately known across the globe as surfing, is used in this study to characterize human-environmental interactions. Although surfing has been an

integral part of Hawaiian identity for millenia[16,17], it has never been used to interpret the relationship between humans and the sea. Next generation 16S rRNA sequencing and liquid chromatography tandem mass spectrometry (LC-MS/MS) technology can be used to describe human environment relationships[18–21] and is used in the SBP to detect environmental signatures in humans.

The surfer biome project sampled an international cohort of 45 surfers from 6 cities in the US, Europe and Africa. In this study, 9 women, 35 men and 1 infant between ages 2 and 48 were investigated. 534 samples from the skin, ear, feces and oral cavity were collected and subjected to LC-MS/MS and 16S rRNA gene sequencing. The samples were processed and implemented into Global Natural Products Social Molecular Networking[22] (GNPS) and Qiita for characterization. Reference microbiome datasets from the American Gut Project[23] (AGP) were used to to identify microbial uniqueness between surfers and non surfers. Similarly, the Earth Microbiome Project[23,24] (EMP) was used as a reference dataset to identify microbial signatures on surfers which were previously isolated from marine environments. Lastly, MASSIVE skin datasets were used to compare metabolomic uniqueness between surfers and non surfers. By investigating the microbiome and metabolome of surfers, this study inclusively illustrates the molecular connection which humans share with their environment.

4.2 Methods

4.2.1 Location selection

Sites were selected to represent a diverse range of regions which hold ocean surfers. An organic farming surf community from Hoy Farms in Lahinch, Ireland was the first selected site. Interestingly London, England has an emerging surf culture and the Backwash Surf magazine launch party at Allpress coffee house selected to represent the non-coastal surf community. Surf vacation hotel, Surf Maroc in Agadir Morocco were selected as the third surf location. The Scripps Institution of Oceanography was selected as the region for San Diego. Ocean Beach San Francisco was selected as a Northern California surf site. Hilo, Hawaii was the final surf location selected in the experiment.

4.2.2 Volunteers/IRB protocol numbers

Forty five volunteers were recruited to participate in this study. All Individuals signed a written informed consent and approved sample collection form their hands and personal objects in accordance with sampling procedures approved by the University of California San Diego Institutional Review Board protocol (IRB number **130537X**).

4.2.3 Sampling

Volunteers were sampled 10 times across their body by taking cotton swabs and pressing to skin and surfboard. Each site on volunteer and sampling location was sampled twice for metabolite profiling and 16s microbial analysis. Sampling was performed at a 2×2 cm area of the forehead, adjacent to tear duct, nostril, tongue, chest, naval, hand, foot and surfboard for each volunteer. Cotton swabs were taken,

sample locations were noted, and swabs were placed in a deep-well 2-mL polypropylene 96-well micro-titer plate.

4.2.4 Extraction

The absorbed material was then extracted in 500µL of 50:50 ethanol/water (for mass spectrometric analysis) or Tris-EDTA buffer (50 mM Tris pH 7.6, 1 mM EDTA, and 0.5% Tween 20) for bacterial DNA extraction. After DNA extraction, amplification and sequencing according to the 16S rRNA gene amplicon protocol of the Earth Microbiome Project v.4.13 (http://earthmicrobiomeproject.org) was performed[25]. DNA was extracted, amplified and sequenced at UC San Diego according to the 16S rRNA amplicon protocol of the Earth Microbiome Project v.4.13 (http://earthmicrobiomeproject.org).

4.2.5 UPLC-MS/MS

Processed office/volunteer extracts (5uL) were subjected to UHPLC chromatographic separation using an UltiMate 3000 UHPLC system (Thermo Scientific), controlled by Chromeleon software (Thermo Scientific). Chromatographic separation was achieved using an 1.7 micron C18 (50 x 2.1 mm) Kinetex UHPLC column (Phenomenex) at 40°C, using a flow rate of 0.5 mL/min. A linear gradient was used for the separation: 0-0.5 min 5% B, 0.5-8 min 5%B-99%B, 8-9 min 99%B, 9.01-10 min 1%B, 10-10.5 min 5%-99% B, 11-11.5 min 99%B-1%B, 12-12.5 min 1% B where solvent A is water 0.1% formic acid (v/v) and solvent B is acetonitrile with 0.1% formic acid (v/v). Column eluent was introduced directly into a Bruker Daltonics maXis Impact quadrupole-time-of-flight mass spectrometer equipped with an Apollo II electrospray ionization source and controlled via otofControl v3.4 (build 16) and Hystar v3.2 software

packages (Bruker Daltonics). The maXis instrument was first externally calibrated using ESI-L Low Concentration Tuning Mix (Agilent Technologies) prior to initiation of the sequence of samples, and hexakis (1H,1H,2H-difluoroethoxy)phosphazene (Synquest Laboratories), m/z 622.0295089613, was continuously introduced as an internal calibrant (lock mass) during the entirety of each LC/MS run. Data was collected in positive ion mode, scanning from 80-2000 m/z. Instrument source parameters were set as follows: nebulizer gas (Nitrogen) pressure, 2 Bar; Capillary voltage, 4,500 V; ion source temperature, 200°C; dry gas flow, 9 L/min. MS1 spectral acquisition rate was set at 3Hz and MS/MS acquisition rate was variable (5-10Hz) depending on precursor intensity. Data-dependent MS/MS acquisition was programmed to the top five most intense precursors per MS scan and any precursor was actively excluded for 1 minute after being fragmented twice. Each MS/MS scan acquired was the average of 4 collision energies, paired optimally with specific collision RF (or “ion cooler RF”) voltages and transfer times in order to maximize the qualitative structural information from each precursor. The auto-acquisition of MS/MS spectra was carried according to specific settings. Precursor *m/z* 100, 300, 500, 1000 was selected; with an isolation width of 2, 4, 6, 8; with a base collision energy (eV) of 10, 25, 30, 50; a sampled collision energy of (5, 10, 15, 20), (12.5, 25, 37.5, 48), (15, 30, 45, 60), (25, 40, 75, 100); a collision RF (Vpp) of 250, 500, 1000, 1500 for all precursors; and a transfer time (µsec) of 50, 75, 100, 150 for all samples was used respectively. Data is publically available at http://gnps.ucsd.edu under accession numbers MSV000080561, MSV000080444, MSV000080443 ,MSV000080442.

4.2.6 Feature finding

MZmine 2.0[26] was used to identify molecular features from processed feature XML format. Unique features, represented by MS1 molecular mass, were identified based on detected precursor mass and retention time. MS1 peak detection threshold was set at 5E3 and MS2 peak detection threshold was set at 1E2. Chromatogram filter threshold was set at 5E3 with a minimum peak height threshold set at 1E3. Peak duration filter was set between 0-2min and baseline threshold was set at 1E2. Mz scanning range window was set at 0.02 daltons and the RT scanning range was set at 0.2 min. Deisotoping filter was performed at 20 ppm error and 0.2 RT with a max charge of +4. Similar peaks within a 20 ppm m/z error and 0.2 RT window were consolidated with 0.75 weight on m/z similarity and 0.25 weight on RT similarity. Gap filling was performed at 10% with peaks within 0.1 min RT. An intensity table relating abundances of molecular feature per sample site was obtained automatically from samples. Files were exported as .csv files. To avoid batch effects post data acquisition, control blank molecular features were removed from the dataset unless molecular features were found to be at least 10x more abundant in experimental samples.

4.2.7 Sequencing

DNA extraction and V4 paired end sequencing from stool were performed according to the Earth Microbiome Project Protocols, as previously described and sequenced using an Illumina MiSeq (La Jolla, CA). Processed tables can be found in https://qiita.ucsd.edu/study/description/10244, in addition sequences can be found in EBI under accession number ERP020615.

Amplicon sequences were demultiplexed and quality controlled using the defaults as provided by QIIME 1.9.1. The primary OTU table was generated using Qiita (https://qiita.ucsd.edu/), which uses deblurring. Sequences from EMP, AGP and SMP were trimmed to 90 read and bloom filtering was applied for multivariate analysis.

4.2.8 QIIME

QIIME 1.9.1 was used to compute Bray distances on 16S datasets between SBP and AGP. The distance matrices were partitioned by sample type. Principal coordinates were computed per sample type.

4.2.9 Statistical visualization

MS Principal coordinate analysis was performed using a inhouse Clusterapp (http://137.110.133.15:3838/clusterMetaboApp0.9.1/). Dataset was normalized with a probabilistic quotient normalization parameter at 30%. Bray Curtis Faith distance method was used with a HCA grouping method. PCoA plots were exported a emperor plots[27]. 16 S rRNA PCoA plots were creating using Qiita workflow (qiita.ucsd.edu/study/description/11404).

4.2.10 GNPS molecular networking

Global Natural Products Social Molecular Networking (http://gnps.ucsd.edu) was used to identify MS/MS spectral matches of molecular features in each sample set. Parameters for creating molecular network include a parent mass tolerance=0.1 Da, Min matched Peak=6, Ion Tolerance=0.5, Score Threshold=0.80 with default advance and filter search options (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=87761b36c8ad473ea104aaf971b51

5fa). Using these parameters, 762792 MS/MS spectra were merged into 23223 nodes. Raw data were uploaded to http://gnps.ucsd.edu and cataloged with a massive IDs of MSV000080561, MSV000080444, MSV000080443 ,MSV00008044. Networking parameters can be found with GNPS ID: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=87761b36c8ad473ea104aaf971b515fa. GNPS False discovery rates were applied (ID=974308943a8d48ceb5643ae42aea657d**).** Cystoscape 3.1 was used to visualize output data from GNPS molecular networking.

4.2.11 Qiita

Qiita was used to compute Bray distances between 16S and MS SBP data. The distances were partitioned by geographic region. Taxonomy of 16S were assigned in Qiita as well. Qiita was also used to calculate MS and 16S shannon diversity of SBP data using a Kruskal-wallis test.

4.3 Results

Though environmental bacteria was expected to be found in surfer samples, it was unclear to which sample locations on the human body would possess the most marine microbial signatures. To investigate this, 8500 EMP water (saline) sOTUs were used as reference datasets to determine if there was microbial overlap between 534 surfer body sites and environmental bacteria. Of the 8500 EMP water (saline) sOTUs used in the reference test, 1244 sOTUs (~15%) of the surveyed EMP marine samples were detected in surfer samples (Figure 1a). Of the shared samples, 988 of the EMP marine sequences were detected at least once in skin, 986 were detected at least once

in the medial canthus, 778 were detected at least once in nose, 950 detected at least once in the ear, 422 detected at least once in the feces, and 336 were found at least once in the oral cavity (Figure 1a).

Principal coordinate analysis was performed to characterize the relationship between surfers and non-surfers. Samples from the American Gut Project were used as comparative data sets. Non-surfer AGP samples from the hand, feces and forehead showed significance microbial difference across PCoA space when compared to surfer samples of the surfer biome project (Figure 1b,c,d). No microbial differences were identified between the oral, medial canthus and eye samples. Similarly, molecular uniqueness of surfers and non-surfers were compared across the metabolome by comparing SBP surfers with non-surfer counterparts. Samples from San Diego and Amazonian skin studies were obtained from the University of California's public mass spectrometry repository MassIVE and used as comparative datasets. Significant variation between the groups were observed (Figure 1e).

Within the surfing population, molecular and microbial diversity were investigated based on country of residence. Kruskal-wallis tests were performed to look at shannon index diversity across Moroccan, American, and European surfers. All surfers from each country showed to have significantly different microbial diversity with Irish surfer having the highest diversity and US surfers having the lowest (Figure 2a). Similar metabolomic diversity results were observed with Ireland surfers having the highest molecular diversity and US surfers having the lowest (Figure 2b). Geographic variance was also

observed to be significantly different between the microbial and metabolomic PCoA space (Figure 2c,d).

When manually investigating how participant behavior can affect the surfer biome, increased sunscreen use seen to be correlated with a decrease in skin microbial diversity (Figure 2a). GNPS molecular networking yielded 24,050 nodes that represent individual molecules within the surfer biome (Figure 2b). Approximately 6% of these nodes were Metabolomics Standard Initiative's level 3 annotated[28]. Within the metabolome, UV absorber and key chemical ingredient of sunscreen, octocrylene was identified. This was also found to be in a majority of the sample files from the surfer biomes (12%) status link (https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=e60bfa232bce43c198fc241456252af7&view=view_compound_occurrence#%7B%22table_sort_history%22%3A%22main.TotalFiles_dsc%22%7D).

To classify the variance between the amount of time spent in the ocean (surf frequency) across the surfer microbiome, Random Forest analysis was performed. Both *Psychrobacter spp* and *Pseudoalteromonas spp* were marine bacteria identified among other known human skin associated bacteria to describe microbial variance (Figure 4a). When profiling the surfer skin microbiome, it was observed that with increased frequency of surf, there is a significantly greater abundance of *Psychrobacter spp.* present (Figure 4b). To investigate the spatial abundance of *Psychrobacter and Pseudoalteromonas spp* across the US, AGP participant sOTUs were profiled and compared to common skin bacteria such as *Propionibacterium spp*. It was observed

that spatial distribution of both *Psychrobacter* and *Pseudoalteromonas* are most abundant near coastal regions (Figure 4c,e). *Propionibacterium spp* do not appear to have the same relative abundance (Figure 3d) and are found ubiquitously abundant throughout the sampled US.

4.4 Discussion

At the foundation of Hawaiian culture, there lies a fundamental respect and appreciation for the natural world. This perspective lends itself especially advantageous when attempting to explain the relationship that humans and nature. Native Hawaiian *olelo no'eau*, or traditional proverbs, elude to the importance of observation. *I ka nana no, a ike* is one such *olelo no'eau* that emphasises how knowledge is acquired through observation. Subsequently, the Hawaiian saying, *nana i ka ili* suggests that similarities and differences among people and place can be identified by "looking to the skin(Nawahi 1912)." A result clearly exemplified in the SBP. A majority of the EMP marine bacteria sOTUs were found to be shared with the skin (Figure 1a), and it was in fact the human skin that help classify surfer uniqueness among larger populations.

Although many studies across Hawaiian literature explain the indigenous belief that a human's connection to earth transcends physical realities(Kanaiaupuni and Malone 2006), this study offers complementary evidence that describes how geography can affect biological uniqueness. Interestingly, surfers from Ireland appeared to have the highest molecular and metabolomic diversity. In addition to being avid surfers, the participants from Ireland are all organic farmers spending upwards of 80 hours a week

tending the land. Their amphibious-like immersion in both soil and sea may offer new insights into how marine and terrestrial natural spaces can affect a human biome.

Similar to how spending more time in different environmental spaces can affect skin microbial diversity, adding chemicals such as octocrylene to skin seems to affect surfer biome diversity as well. It was interesting to correlate the increase use of sunscreen among surfers to a lesser microbial diversity on the skin and that correlations between cosmetic use and varied skin microbial diversity were not observed (supplemental figure 1). This finding illustrates the added value of characterizing the surfer metabolome alongside microbiome analysis and helps to address its role the metabolome has in facilitating biome diversity[29].

An important finding of the SBP was the ability to detect marine associated bacteria *Psychrobacter spp*[29–32] and *Pseudoalteromonas spp*[33–35] in surfers. The dose-dependent response that relates *Psychrobacter* abundance to surfing frequency and the higher presence in coastal US communities (Figure 4c,d,e) suggest these organisms as potential biomarkers that describes human contact with the ocean.

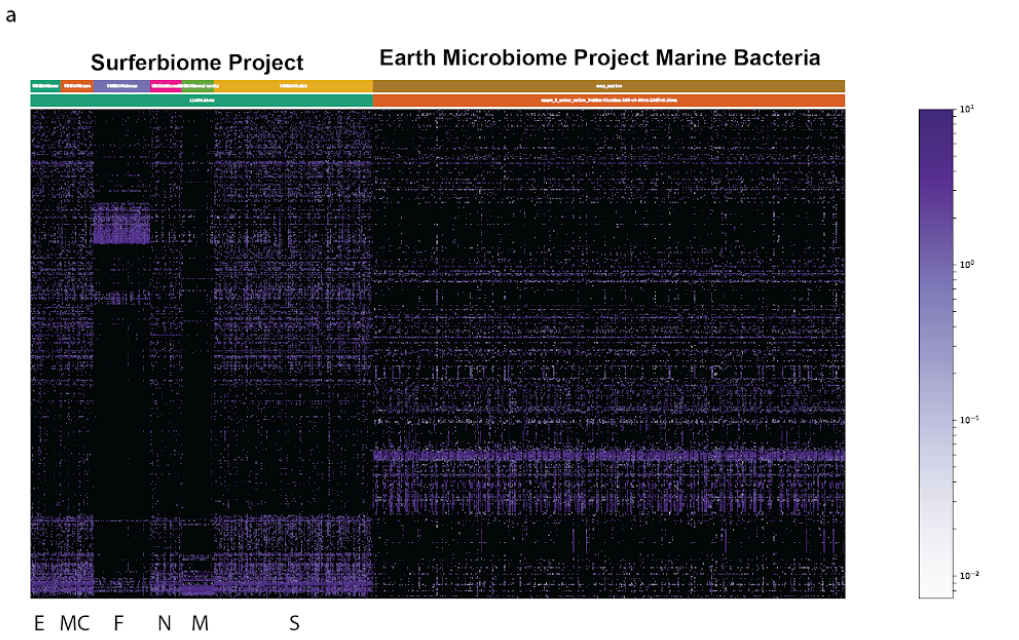
Although one would expect the highly abundant oceanic SAR11 bacteria clade[36] or the effective marine surface colonizer rhodobacter[37,38] to be the most abundant marine bacteria on surfers, it was *psychrobacter* that lends itself as the primary descriptor of ocean interaction. In addition to *Psychrobacter* bacteria being observed across many marine hosts[39–41] this bacteria has also been found on the skin of whales[42–44]. Ancient Hawaiian beliefs have long described the spiritual relationships that humans have with cetaceans and that they have also been described as being

kinolau(Blaisdell 2005; Brown 2016; Marzan 2016),or alternative forms of the ocean god *Lono*(Lebo 2010). In efforts to aid in the expression of the microbial world from an indigenous perspective, microbes have recently been described as analogous to native spirits[45]. Their presence on the skin of ocean surfers and cetacean suggests their potential to be viewed as *kinolau* of *Kanaloa* in a Hawaiian context.

In conclusion, studies like the SBP help bring empirical support for the traditional knowledge of native peoples and more importantly provide added insights into the characterization of human-environment interactions. This project aims to be a contributing source of information that can be used to improve health and environmental resource management systems across the globe.

4.5 Appendix of figures for chapter 4

Figure 1. Microbial uniqueness of surfers, non-surfers and the surrounding environment. a) Heat map comparing sOTUs of marine bacteria from Earth Microbiome Project (EMP) and sOTUs of surfer ear (E), medial canthus (MC), feces (F), nose (N), mouth (M) and skin (S) in the Surfer Biome Project (SBP). b) Principal coordinate analysis plot (PCoA) illustrating microbial variance between SBP and American Gut Project (AGP) non-surfer hand samples. c) PCoA plot illustrating microbial variance between SBP and AGP non-surfer forehead samples. d) PCoA plot illustrating microbial variance between SBP and AGP non-surfer fecal samples. e) PCoA plot illustrating metabolome variance between SBP and AGP non-surfer of hand samples.


a
Surferbiome Project
Earth Microbiome Project Marine Bacteria
E MC F N M S

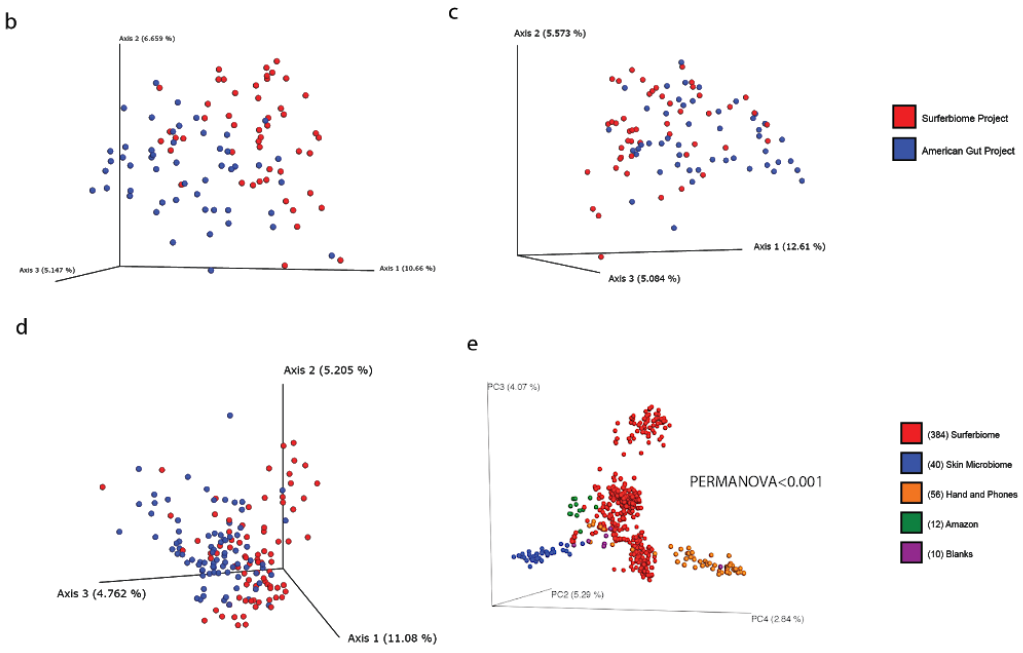

b
Axis 2 (6.659 %)
Axis 3 (5.147 %)
Axis 1 (10.66 %)
c
Axis 2 (5.573 %)
Axis 1 (12.61 %)
Axis 3 (5.084 %)
Surferbiome Project
American Gut Project
d
Axis 2 (5.205 %)
Axis 3 (4.762 %)
Axis 1 (11.08 %)
e
PC3 (4.07 %)
PERMANOVA<0.001
PC2 (5.29 %)
PC4 (2.84 %)
(384) Surferbiome
(40) Skin Microbiome
(56) Hand and Phones
(12) Amazon
(10) Blanks

Figure 2. Metabolomic and microbial shannon diversity analysis. a) Kruskal-wallis box plot of metabolomic diversity across Morocco, San Diego, San Francisco, Hilo, London, and County Clare. b) Kruskal-wallis box plot of microbial diversity across Morocco, San Diego, San Francisco, Hilo, London, and County Clare. c) PCoA bray curtis distance method analysis of metabolomic diversity across Morocco, San Diego, San Francisco, Hilo, London, and County Clare d) PCoA bray curtis distance method analysis of microbial diversity across Morocco, San Diego, San Francisco, Hilo, London, and County Clare

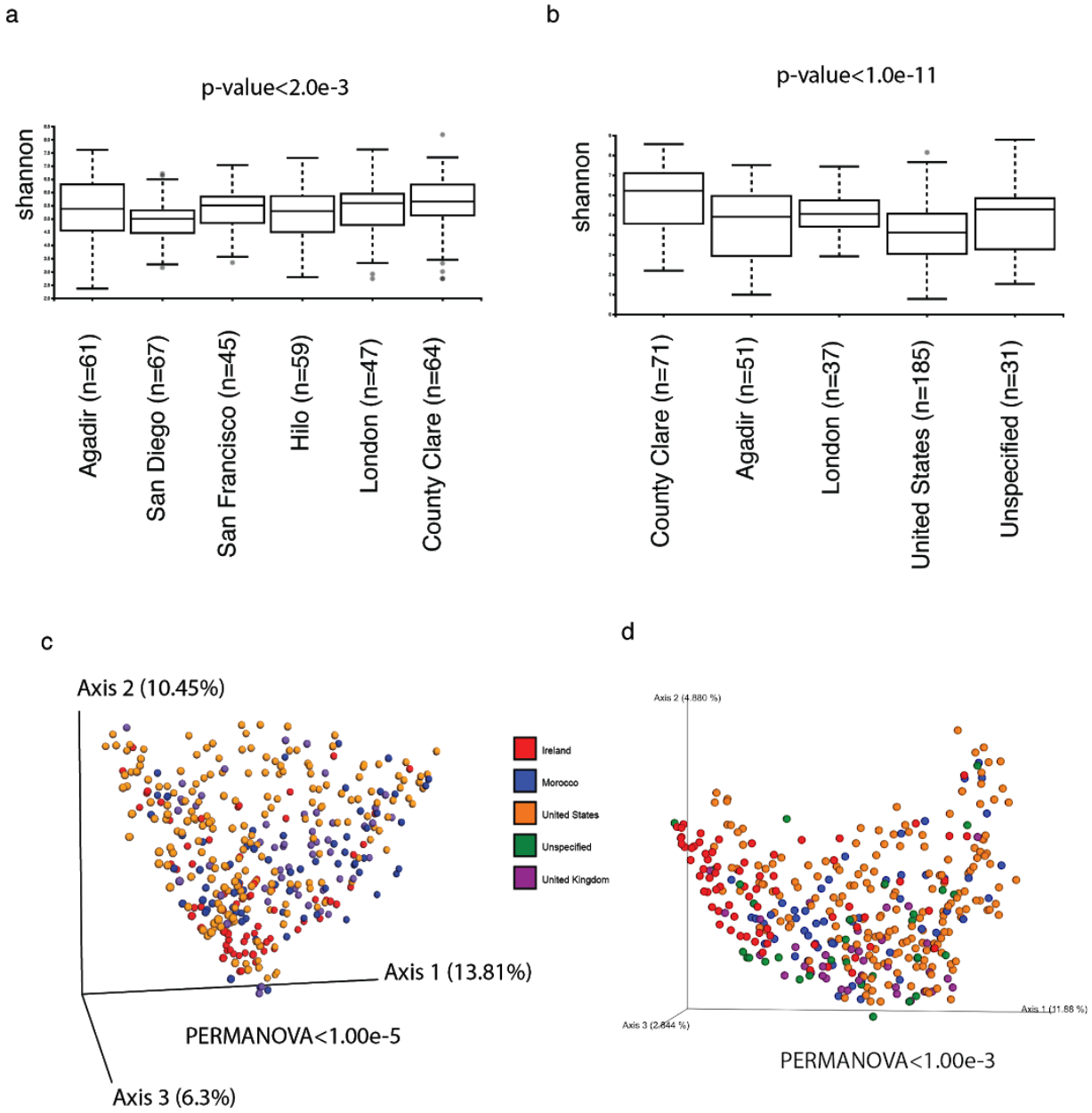

a
p-value<2.0e-3
shannon
Agadir (n=61)
San Diego (n=67)
San Francisco (n=45)
Hilo (n=59)
London (n=47)
County Clare (n=64)
b
p-value<1.0e-11
shannon
County Clare (n=71)
Agadir (n=51)
London (n=37)
United States (n=185)
Unspecified (n=31)
c
Axis 2 (10.45%)
Axis 1 (13.81%)
Axis 3 (6.3%)
PERMANOVA<1.00e-5
Ireland
Morocco
United States
Unspecified
United Kingdom
d
PERMANOVA<1.00e-3

Figure 3. Sunscreen use frequency and the effects on microbial diversity of skin samples from SBP. Kruskal-wallis box plot of microbiome diversity in surfers in relation to frequency of sunscreen use. b) GNPS molecular network of known molecules highlighted by boxes. Gray nodes represent unannotated nodes and light blue nodes represent molecules identified via GNPS molecular networking. c) Molecular cluster with matches to molecular feature 362.212 m/z octocrylene.

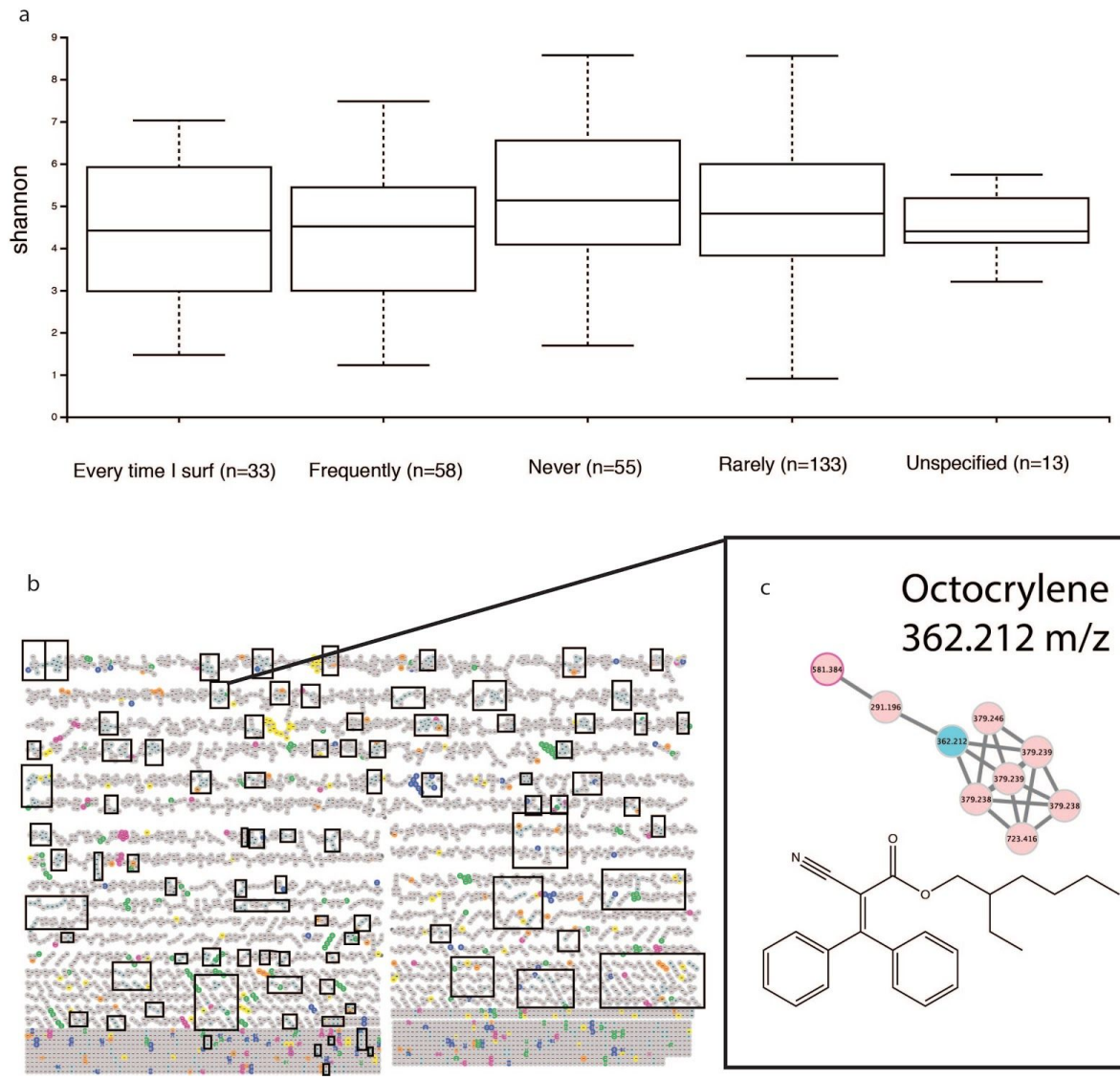
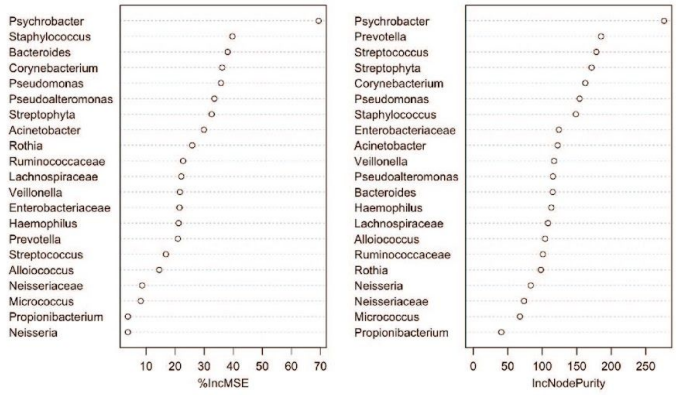
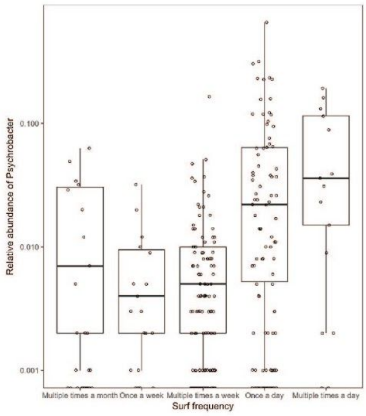

a
shannon
0
1
2
3
4
5
6
7
8
9
Every time I surf (n=33)
Frequently (n=58)
Never (n=55)
Rarely (n=133)
Unspecified (n=13)
b
c
Octocrylene
362.212 m/z
581.384
291.196
362.212
379.246
379.239
379.239
379.238
379.238
723.416
N
O
O

Figure 4. *Psychrobacter spp* as a molecular describer of ocean interaction. a) Random Forest analysis classifying the top 20 bacteria that describe surfing frequency. b) Box plot describing the relative abundance of *Psychrobacter spp* as surfing frequency increases. c) Spatial mapping of *Psychrobacter spp* abundance in the USA. d) Spatial mapping of *Propionibacterium spp* abundance in the USA. e) Spatial mapping of *Pseudoalteromonas spp* abundance in the USA.

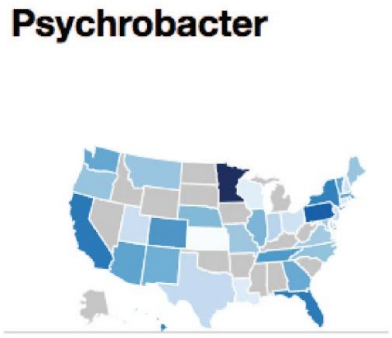
a


Psychrobacter
Staphylococcus
Bacteroides
Corynebacterium
Pseudomonas
Pseudoalteromonas
Streptophyta
Acinetobacter
Rothia
Ruminococcaceae
Lachnospiraceae
Veillonella
Enterobacteriaceae
Haemophilus
Prevotella
Streptococcus
Alloiococcus
Neisseriaceae
Micrococcus
Propionibacterium
Neisseria
10 20 30 40 50 60 70
%IncMSE
Psychrobacter
Prevotella
Streptococcus
Streptophyta
Corynebacterium
Pseudomonas
Staphylococcus
Enterobacteriaceae
Acinetobacter
Veillonella
Pseudoalteromonas
Bacteroides
Haemophilus
Lachnospiraceae
Alloiococcus
Ruminococcaceae
Rothia
Neisseria
Neisseriaceae
Micrococcus
Propionibacterium
0 50 100 150 200 250
IncNodePurity

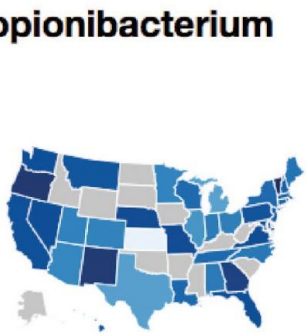
b


Relative abundance of Psychrobacter
0.100
0.010
0.001
Multiple times a month
Once a week
Multiple times a week
Once a day
Multiple times a day
Surf frequency

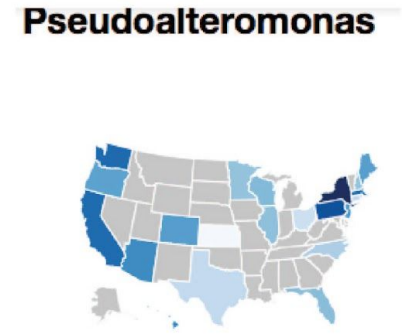
c


Psychrobacter

d


Propionibacterium

e


Pseudoalteromonas

4.6 Acknowledgments

The work presented in Chapter 4 is the result of collaborations with the Knight lab at the University of California San Diego. Dorrestein lab member, Mike Meehan provided assistance in the overnight shipping of surfer samples after collection. Knight lab member, Lindsay DeRight Goldasich performed DNA extractions and prepared them for Illumina sequencing. Gail Ackerman provided assistance during upload of both the microbiome and metabolome metadata in Qiita. Knight lab members Jon Sanders and Anupriya Tripathi created Figure 1a comparing the sOTUs data from SBP and EMP. Knight lab member Chris Callewaert created Figure 4a and 4b which identifies the relationship between Psychrobacter and surf frequency in SBP dataset. Daniel Mcdonald compared the 16S microbiome sOTUs of the AGP body site samples with SBP body sites as seen Figure 1b, 1c, and 1d. Dorrestein lab member Ricardo Silva offered insight into statistical analysis techniques. Dorrestein lab member Louis Felix and Madeleine Ernst offered insight into molecular networking annotation. Dorrestein lab member Daniel Petras provided insight into MZmine2 data processing techniques for MS feature detection. Chapter 4 is currently being prepared for submission of the material. Kapono, C.A., Tripathi, A., Callewaert, C., Mcdonald, D., Sanders, J., Da Silva, R., Nothias, L.F., Ernst, M., Petras, D., Deright Goldasich, L.D., Akermann, G., Knight, R., Dorrestein, P.C. The dissertation author was the primary investigator and author of this manuscript.

4.7 References

1. Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP, Bohannan BJM. Identifying personal microbiomes using metagenomic codes. Proc Natl Acad Sci U S A. 2015 Jun 2;112(22):E2930–8.

2. NIH HMP Working Group, Peterson J, Garges S, Giovanni M, McInnes P, Wang L. The NIH Human Microbiome Project. Genome Res. 2009 Dec;19(12):2317–23.

3. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary microbiome. Genome Res. 2009 Apr;19(4):636–43.

4. Oh J, Byrd AL, Park M, NISC Comparative Sequencing Program, Kong HH, Segre JA. Temporal Stability of the Human Skin Microbiome. Cell. 2016 May 5;165(4):854–66.

5. Oh J, Byrd AL, Deming C, Conlan S, NISC Comparative Sequencing Program, Kong HH. Biogeography and individuality shape function in the human skin metagenome. Nature. 2014 Oct 2;514(7520):59–64.

6. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. Nature. 2011 Oct 30;480(7376):241–4.

7. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M. Human gut microbiome viewed across age and geography. Nature. 2012 May 9;486(7402):222–7.

8. Pierotti R, Wildcat D. Traditional Ecological Knowledge: The Third Alternative (Commentary). Ecol Appl. 2000;10(5):1333.

9. Rist S, Dahdouh-Guebas F. Ethnosciences—A step towards the integration of scientific and indigenous forms of knowledge in the management of natural resources for the future. Environ Dev Sustainability. 2006;8(4):467–93.

10. Aikenhead GS, Ogawa M. Indigenous knowledge and science revisited. Cult Stud Sci Educ. 2007;2(3):539–620.

11. Furet J-P, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. FEMS Microbiol Ecol. 2009 Jun;68(3):351–62.

12. Fujimura KE, Demoor T, Rauch M, Faruqi AA, Jang S, Johnson CC. House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. Proc Natl Acad Sci U S A. 2014 Jan

14;111(2):805–10.

13. Shahid S, Finn L, Bessarab D, Thompson SC. Understanding, beliefs and perspectives of Aboriginal people in Western Australia about cancer and its impact on access to cancer services. BMC Health Serv Res [Internet]. 2009;9(1). Available from: http://dx.doi.org/10.1186/1472-6963-9-132

14. Walker DH, Sinclair FL, Thapa B. Incorporation of indigenous knowledge and perspectives in agroforestry development. Agrofor Syst. 1995;30(1-2):235–48.

15. Broadhead L-A, Howard S. Deepening the debate over “sustainable science”: Indigenous perspectives as a guide on the journey. Sustainable Development. 2009;19(5):301–11.

16. Walker IH. Waves of Resistance: Surfing and History in Twentieth-century Hawai‘i. University of Hawaii Press; 2011. 225 p.

17. Finney BR, Houston JD. Surfing: A History of the Ancient Hawaiian Sport. Pomegranate; 1996. 117 p.

18. Kapono CA, Morton JT, Bouslimani A, Melnik AV, Orlinsky K, Knaan TL. Creating a 3D microbial and chemical snapshot of a human habitat. Sci Rep. 2018 Feb 27;8(1):3669.

19. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011 Apr;9(4):279–90.

20. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014 Aug 29;345(6200):1048–52.

21. Hanski I, von Hertzen L, Fyhrquist N, Koskinen K, Torppa K, Laatikainen T. Environmental biodiversity, human microbiota, and allergy are interrelated. Proc Natl Acad Sci U S A. 2012 May 22;109(21):8334–9.

22. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016 Aug 9;34(8):828–37.

23. Debelius JW, Xu Z, Vázquez-Baeza Y, Knight R, Wolfe E, McDonald D. Turning Participatory Microbiome Research into Usable Data: Lessons from the American Gut Project. J Microbiol Biol Educ. 2016;17(1):46–50.

24. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ. A communal catalogue reveals Earth’s multiscale microbial diversity. Nature. 2017 Nov 23;551(7681):457–63.

25. Gilbert J. The Earth Microbiome Project: A new paradigm in geospatial and

temporal studies of microbial ecology [Internet]. SciVee. 2012. Available from: http://dx.doi.org/10.4016/46411.01

26. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11(1):395.

27. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience. 2013 Nov 26;2(1):16.

28. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007 Sep;3(3):211–21.

29. Bozal N, Montes MJ, Tudela E, Guinea J. Characterization of several Psychrobacter strains isolated from Antarctic environments and description of Psychrobacter luti sp. nov. and Psychrobacter fozii sp. nov. Int J Syst Evol Microbiol. 2003 Jul;53(Pt 4):1093–100.

30. Bowman JP, Nichols DS, McMeekin TA. Psychrobacter glacincola sp. nov., a Halotolerant, Psychrophilic Bacterium Isolated from Antarctic Sea Ice. Syst Appl Microbiol. 1997;20(2):209–15.

31. Bowman JP, Cavanagh J, Austin JJ, Sanderson K. Novel Psychrobacter species from Antarctic ornithogenic soils. Int J Syst Bacteriol. 1996 Oct;46(4):841–8.

32. Ponder MA, Gilmour SJ, Bergholz PW, Mindock CA, Hollingsworth R, Thomashow MF. Characterization of potential stress responses in ancient Siberian permafrost psychroactive bacteria. FEMS Microbiol Ecol. 2005 Jun 1;53(1):103–15.

33. Holmstram C. Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. FEMS Microbiol Ecol. 1999;30(4):285–93.

34. El Gamal A, Agarwal V, Diethelm S, Rahman I, Schorn MA, Sneed JM. Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase–thioesterase enzyme pair. Proceedings of the National Academy of Sciences. 2016;113(14):3797–802.

35. Marsh G, Athanasiadou M, Athanassiadis I, Bergman A, Endo T, Haraguchi K. Identification, quantification, and synthesis of a novel dimethoxylated polybrominated biphenyl in marine mammals caught off the coast of Japan. Environ Sci Technol. 2005 Nov 15;39(22):8684–90.

36. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA. SAR11 clade dominates ocean surface bacterioplankton communities. Nature.

2002;420(6917):806–10.

37. Crump BC, Armbrust EV, Baross JA. Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. Appl Environ Microbiol. 1999 Jul;65(7):3192–204.

38. Dang H, Li T, Chen M, Huang G. Cross-ocean distribution of Rhodobacterales bacteria as primary surface colonizers in temperate coastal marine waters. Appl Environ Microbiol. 2008 Jan;74(1):52–60.

39. Ringø E, Sperstad S, Myklebust R, Refstie S, Krogdahl Å. Characterisation of the microbiota associated with intestine of Atlantic cod (Gadus morhua L.). Aquaculture. 2006;261(3):829–41.

40. Deschaght P, Janssens M, Vaneechoutte M, Wauters G. Psychrobacter isolates of human origin, other than Psychrobacter phenylpyruvicus, are predominantly Psychrobacter faecalis and Psychrobacter pulmonis, with emended description of P. faecalis. Int J Syst Evol Microbiol. 2012 Mar;62(Pt 3):671–4.

41. Mellish J, Tuomi P, Hindle A, Jang S, Horning M. Skin microbial flora and effectiveness of aseptic technique for deep muscle biopsies in Weddell seals (Leptonychotes weddellii) in McMurdo Sound, Antarctica. J Wildl Dis. 2010 Apr;46(2):655–8.

42. Bierlich KC, Miller C, DeForce E, Friedlaender AS, Johnston DW, Apprill A. Temporal and Regional Variability in the Skin Microbiome of Humpback Whales along the Western Antarctic Peninsula. Appl Environ Microbiol [Internet]. 2018 Mar 1;84(5). Available from: http://dx.doi.org/10.1128/AEM.02574-17

43. Nelson TM, Apprill A, Mann J, Rogers TL, Brown MV. The marine mammal microbiome: current knowledge and future directions. Microbiol Aust. 2015;36(1):8.

44. Apprill A, Robbins J, Murat Eren A, Pack AA, Reveillaud J, Mattila D. Humpback Whale Populations Share a Core Skin Bacterial Community: Towards a Health Index for Marine Mammals? PLoS One. 2014;9(3):e90785.

45. Giraldo Herrera CE. Microbes and Other Shamanic Beings. 2018.

5. Chapter 5

5.1 Future Directions

From identifying new forms of “molecular fingerprints” (chapter 2), to spatially mapping coral interactions (chapter 3), to discovering the marine signatures found on surfers (chapter 4), multi-omic molecular profiling has provided me with new and exciting ways to interpret the natural world. A perspective that I would like to continue to develop after the completion of this program especially when promoting and advancing science research in Hawaii.

Research in chapter 2 has laid the preliminary groundwork in connecting organisms to place. Often times invasive species in Hawaii become suspects of ecological disruption[1–4], but very little evidence exists that traces behavior back to individuals. By integrating molecular profiling as a forensic application to environmental management studies, investigators may be able to interpret and better manage invasive species across the islands. One potential example of this type of study could include the spatial mapping of *Ceratocystis* fungus and its carriers across healthy and diseased *Metrosideros polymorpha* trees[5]. *Ceratocystis* infection has recently plagued the forest of Hawaii causing massive death to the native tree species *M. polymorpha*. Having significant ecological[6–8] and cultural value, new research efforts look to identify potential vectors of transmission. Many of the affected trees in the forest have been mapped across the island[9], but only one species of burrowing beetle has be listed as a potential suspect of transmission. Potential applications for research described in chapter 2 could entail the molecular profiling of both tree as well as known organisms that interact with

the tree such as insects, birds, rodents and humans. Characterizing microbial and chemical relationship between these forest organisms and their “work” environment may offer new insights into the potential of disease transmission. Additionally, by combining both targeted and untargeted molecular profiling, known carriers of *Ceratocystis* might be rapidly identified.

As coral reef health continues to be a concerning issues across Hawaii[10–15], research performed in chapter 3 may provide coral researchers with new tools when characterizing complex coral assemblages. Recent interests in furthering coral research across the main eight Hawaiian islands and into the Papahanaumokuakea monument[16] may benefit from the integration of 3d molecular profiling. Several research labs in Hawaii including the Burns, Gates and Takabayashi lab are applying novel 3d mapping technology to coral communities across the Northwestern Hawaiian Islands. Taking inventories of microbial and metabolomic features across coral assemblages and mapping them onto 3d models may provide powerful insights during the analysis of long term ecological studies[17–19].

The surfer biome project described in chapter 4 has gained international attention and also presents an effective avenue to communicate human environmental interactions. A major finding that describes the molecular relationship that humans share with the ocean is the detection of marine bacteria *Psychrobacter* and *Pseudoalteromonas* on surfer skin. It remains unclear whether these bacteria improve human health or even if these organisms are actively colonizing the skin. Future directions of study include further looking at antagonistic interaction between

Psychrobacter and/or *Pseudoalteromonas* and human skin pathogens[20–23]. Although it is not trivial to investigate if environmental marine bacteria colonizes human skin, potential directions of research may include targeted metabolomics for known *Psychrobacter* and *Pseudoalteromonas* metabolites on human skin. Incorporating other data acquisition MS technologies such as GM-MS/MS[24,25] and hybrid quadrupole analyzers[26–29] may increase MS sensitivity and quantitative resolution enough to effectively detect targeted specialized metabolites of marine bacteria colonization.

Lastly, my graduate research has helped me recognize the importance of science communication both within the scientific community as well as across broader audiences. My research has also provided me the ability to gain a voice that extends beyond academia and into different social spaces. As a recent recipient of the American Association for the Advancements of Science 2018 Mass Media fellowship, I will begin formal training as a science journalist. Having been awarded a fellowship, I will begin contributing science t the Union Tribune San Diego starting June 2018. I hope that this opportunity will provide me with the necessary training to improve my scientific writing which I can later use to communicate my future research to broader audiences. Alternatively, this experience will hopefully allow me to explore a career in science writing and better advance science consciousness across broader audiences.

5.3 References

1. Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP, Bohannan BJM. Identifying personal microbiomes using metagenomic codes. Proc Natl Acad Sci U S A. 2015 Jun 2;112(22):E2930–8.
2. NIH HMP Working Group, Peterson J, Garges S, Giovanni M, McInnes P, Wang L.

The NIH Human Microbiome Project. Genome Res. 2009 Dec;19(12):2317–23.

3. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary microbiome. Genome Res. 2009 Apr;19(4):636–43.

4. Oh J, Byrd AL, Park M, NISC Comparative Sequencing Program, Kong HH, Segre JA. Temporal Stability of the Human Skin Microbiome. Cell. 2016 May 5;165(4):854–66.

5. Oh J, Byrd AL, Deming C, Conlan S, NISC Comparative Sequencing Program, Kong HH. Biogeography and individuality shape function in the human skin metagenome. Nature. 2014 Oct 2;514(7520):59–64.

6. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. Nature. 2011 Oct 30;480(7376):241–4.

7. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M. Human gut microbiome viewed across age and geography. Nature. 2012 May 9;486(7402):222–7.

8. Pierotti R, Wildcat D. Traditional Ecological Knowledge: The Third Alternative (Commentary). Ecol Appl. 2000;10(5):1333.

9. Rist S, Dahdouh-Guebas F. Ethnosciences—A step towards the integration of scientific and indigenous forms of knowledge in the management of natural resources for the future. Environ Dev Sustainability. 2006;8(4):467–93.

10. Aikenhead GS, Ogawa M. Indigenous knowledge and science revisited. Cult Stud Sci Educ. 2007;2(3):539–620.

11. Furet J-P, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. FEMS Microbiol Ecol. 2009 Jun;68(3):351–62.

12. Fujimura KE, Demoor T, Rauch M, Faruqi AA, Jang S, Johnson CC. House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. Proc Natl Acad Sci U S A. 2014 Jan 14;111(2):805–10.

13. Shahid S, Finn L, Bessarab D, Thompson SC. Understanding, beliefs and perspectives of Aboriginal people in Western Australia about cancer and its impact on access to cancer services. BMC Health Serv Res [Internet]. 2009;9(1). Available from: http://dx.doi.org/10.1186/1472-6963-9-132

14. Walker DH, Sinclair FL, Thapa B. Incorporation of indigenous knowledge and perspectives in agroforestry development. Agrofor Syst. 1995;30(1-2):235–48.

15. Broadhead L-A, Howard S. Deepening the debate over “sustainable science”: Indigenous perspectives as a guide on the journey. Sustainable Development. 2009;19(5):301–11.

16. Walker IH. Waves of Resistance: Surfing and History in Twentieth-century Hawai‘i. University of Hawaii Press; 2011. 225 p.

17. Finney BR, Houston JD. Surfing: A History of the Ancient Hawaiian Sport. Pomegranate; 1996. 117 p.

18. Kapono CA, Morton JT, Bouslimani A, Melnik AV, Orlinsky K, Knaan TL. Creating a 3D microbial and chemical snapshot of a human habitat. Sci Rep. 2018 Feb 27;8(1):3669.

19. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011 Apr;9(4):279–90.

20. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014 Aug 29;345(6200):1048–52.

21. Hanski I, von Hertzen L, Fyhrquist N, Koskinen K, Torppa K, Laatikainen T. Environmental biodiversity, human microbiota, and allergy are interrelated. Proc Natl Acad Sci U S A. 2012 May 22;109(21):8334–9.

22. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016 Aug 9;34(8):828–37.

23. Debelius JW, Xu Z, Vázquez-Baeza Y, Knight R, Wolfe E, McDonald D. Turning Participatory Microbiome Research into Usable Data: Lessons from the American Gut Project. J Microbiol Biol Educ. 2016;17(1):46–50.

24. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ. A communal catalogue reveals Earth’s multiscale microbial diversity. Nature. 2017 Nov 23;551(7681):457–63.

25. Gilbert J. The Earth Microbiome Project: A new paradigm in geospatial and temporal studies of microbial ecology [Internet]. SciVee. 2012. Available from: http://dx.doi.org/10.4016/46411.01

26. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11(1):395.

27. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience. 2013 Nov 26;2(1):16.

28. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007 Sep;3(3):211–21.

29. Bozal N, Montes MJ, Tudela E, Guinea J. Characterization of several Psychrobacter strains isolated from Antarctic environments and description of Psychrobacter luti sp. nov. and Psychrobacter fozii sp. nov. Int J Syst Evol Microbiol. 2003 Jul;53(Pt 4):1093–100.

30. Bowman JP, Nichols DS, McMeekin TA. Psychrobacter glacincola sp. nov., a Halotolerant, Psychrophilic Bacterium Isolated from Antarctic Sea Ice. Syst Appl Microbiol. 1997;20(2):209–15.

31. Bowman JP, Cavanagh J, Austin JJ, Sanderson K. Novel Psychrobacter species from Antarctic ornithogenic soils. Int J Syst Bacteriol. 1996 Oct;46(4):841–8.

32. Ponder MA, Gilmour SJ, Bergholz PW, Mindock CA, Hollingsworth R, Thomashow MF. Characterization of potential stress responses in ancient Siberian permafrost psychroactive bacteria. FEMS Microbiol Ecol. 2005 Jun 1;53(1):103–15.

33. HolmstrÃ¶m C. Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. FEMS Microbiol Ecol. 1999;30(4):285–93.

34. El Gamal A, Agarwal V, Diethelm S, Rahman I, Schorn MA, Sneed JM. Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase–thioesterase enzyme pair. Proceedings of the National Academy of Sciences. 2016;113(14):3797–802.

35. Marsh G, Athanasiadou M, Athanassiadis I, Bergman A, Endo T, Haraguchi K. Identification, quantification, and synthesis of a novel dimethoxylated polybrominated biphenyl in marine mammals caught off the coast of Japan. Environ Sci Technol. 2005 Nov 15;39(22):8684–90.

36. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA. SAR11 clade dominates ocean surface bacterioplankton communities. Nature. 2002;420(6917):806–10.

37. Crump BC, Armbrust EV, Baross JA. Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. Appl Environ Microbiol. 1999 Jul;65(7):3192–204.

38. Dang H, Li T, Chen M, Huang G. Cross-ocean distribution of Rhodobacterales bacteria as primary surface colonizers in temperate coastal marine waters. Appl Environ Microbiol. 2008 Jan;74(1):52–60.

39. Ringø E, Sperstad S, Myklebust R, Refstie S, Krogdahl Å. Characterisation of the microbiota associated with intestine of Atlantic cod (Gadus morhua L.). Aquaculture. 2006;261(3):829–41.

40. Deschaght P, Janssens M, Vaneechoutte M, Wauters G. Psychrobacter isolates of human origin, other than Psychrobacter phenylpyruvicus, are predominantly Psychrobacter faecalis and Psychrobacter pulmonis, with emended description of P. faecalis. Int J Syst Evol Microbiol. 2012 Mar;62(Pt 3):671–4.

41. Mellish J, Tuomi P, Hindle A, Jang S, Horning M. Skin microbial flora and effectiveness of aseptic technique for deep muscle biopsies in Weddell seals (Leptonychotes weddellii) in McMurdo Sound, Antarctica. J Wildl Dis. 2010 Apr;46(2):655–8.

42. Bierlich KC, Miller C, DeForce E, Friedlaender AS, Johnston DW, Apprill A. Temporal and Regional Variability in the Skin Microbiome of Humpback Whales along the Western Antarctic Peninsula. Appl Environ Microbiol [Internet]. 2018 Mar 1;84(5). Available from: http://dx.doi.org/10.1128/AEM.02574-17

43. Nelson TM, Apprill A, Mann J, Rogers TL, Brown MV. The marine mammal microbiome: current knowledge and future directions. Microbiol Aust. 2015;36(1):8.

44. Apprill A, Robbins J, Murat Eren A, Pack AA, Reveillaud J, Mattila D. Humpback Whale Populations Share a Core Skin Bacterial Community: Towards a Health Index for Marine Mammals? PLoS One. 2014;9(3):e90785.

45. Giraldo Herrera CE. Microbes and Other Shamanic Beings. 2018.