UNIVERSITY OF CALIFORNIA SAN DIEGO

The Fundamentals of Tooth Staining

A thesis submitted in partial satisfaction of the

requirements for the degree of Master of Science

in

Biology

by

Kenneth L. Jones II

Committee in charge:

Professor Pieter C. Dorrestein, Chair
Professor Kit Pogliano, Co-Chair
Professor Joe Pogliano
Professor Manuela Raffatellu

2020

The thesis of Kenneth L. Jones II is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2020

DEDICATION

I dedicate this work

to my family who laid down the foundation for me to be here,

my loved ones who have provided unconditional support along the way,

and friends who have kept me smiling.

Thank you to everyone

who has never put a limit to who and what I can be.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ANOISM	Analysis of Similarities
ANOVA	Analysis of Variance
Da	Dalton
GNPS	Global Natural Products Social Molecular Networking
HESI	Heated Electrospray Ionization
Hz	Hertz
ID	Identification
IPA	Isopropanol
IQR	Interquartile Range
LC-MS	Liquid Chromatography Mass Spectrometry
LTQ	Linear Trap Quadrupole
MDA	Mean Decrease Accuracy
MMVEC	Microbe to Metabolite Vectorization
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NAP	Network Annotated Propagation
OOB	Out of the Bag
PLS-DA	Principal Least Square – Discriminant Analysis
PQN	Probabilistic Quotient Normalization
RF	Random Forest
RT	Retention Time
UHPLC	Ultra High-Performance Liquid Chromatography

VIP	Variance of Importance
WIO	Whiteness Index
XL	Extra Large

LIST OF SUPPLEMETNAL FILES

File 1 Rutgers Dataset MMVEC Visualization, Crowded View.qzv

File 2 Therametrics Dataset MMVEC Visualization, Crowded.qzv

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Pieter Dorrestein for his support as the chair of my committee. Through multiple drafts and many long nights, his guidance has proved to be invaluable.

I would also like to acknowledge the Dorrestein Lab, without whom my research would have no doubt taken five times as long. It is their support that helped me in an immeasurable way.

This thesis is coauthored with Melnik, Alexey, Aksenov, Alexander, Minich, Jeremiah, and Lejzerowitz, Franck. The thesis author was the primary author of this material.

VITA

2014 Bachelors of Science, University of California, San Diego

2020 Master of Science, University of California, San Diego

PUBPLICATIONS

"Repository-scale Co- and Re-analysis of Tandem Mass Spectrometry Data" bioRxiv 750471; doi: https://doi.org/10.1101/750471

"Algorithmic Learning for Auto-deconvolution of GC-MS Data to Enable Molecular Networking within GNPS" bioRxiv 2020.01.13.905091; doi: https://doi.org/10.1101/2020.01.13.905091

FIELDS OF STUDY

Major Field: Biology

Studies in Metabolome

Professor Pieter C. Dorrestein

ABSTRACT OF THE THESIS

The Fundamentals of Tooth Staining

by

Kenneth L. Jones II

Master of Science in Biology

University of California, San Diego, 2020

Professor Pieter C. Dorrestein, Chair
Professor Kit Pogliano, Co-Chair

Brilliant white teeth is a desire for most Americans today. Currently there is different whitening products to bring this shine, however the stains themselves do not have structured origins. Throughout their life, a tooth's shade can change due to age, unrelated impacts, a number of oral cavity diseases, and by the products which we consume. Exploring the consumption of our food products, the chemistry of the oral cavity, and the associating microbial communities in relation to tooth staining is the focus of this study. In hopes of determining the chemical compounds that are causing the staining to occur extrinsically (in the enamel) and intrinsically (in the dentin) within the coronal portion of the tooth. In addition to this goal the

study aims to assess the differences between teeth that have been stained significantly over time, versus teeth that have normal natural stain occurrences throughout time. The teeth that have been stained significantly over time were artificially stained with common staining solutions. Those staining solutions include: coffee, tobacco, tea, and wine. These solutions have been index and annotated for comparison to normal naturally stained teeth. 1000 human teeth were used in this study, which were analyzed using HPLC-MS/MS. Based on the metabolomics analysis, microbial communities will be analyzed in correlation to the various staining measurements.

Chapter I

Introduction

A. Importance of Oral Health

The oral cavity is the gateway to understanding a person's overall health history through what it can imply. Oral diseases often have an impact on many aspects of general health including general diseases, which impact the health of the oral cavity. Poor oral health is preventable, however untreated and unresolved oral diseases can have a large impact on a person from the disruption of sleep, to tooth loss, as well as the formation of ulcerations and abscesses [1]. A common disease that intensifies poor oral is a diabetes. Diabetes is a disease that plagues Americans and is known to increase the chances of periodontal disease [2]. Along with periodontal disease, dental caries have become some of the biggest threats to oral health and are a part of the top most chronic diseases in the United States [3]. Both of these diseases vary in their appearance when being discovered within enamel, the outer layer the tooth.

The variation can be related to old age due to age, unrelated impacts, a number of oral cavity diseases, and by the products which we consume [4]. Assessing and discerning the appearance of common disease allows for differences between a healthy and an unhealthy tooth to be identified. This also improves the ability to discern specific staining agents from disease.

B. Disease Related Discoloration

Due to the commonality of dental caries and periodontal disease amongst Americans, it's imperative to define characteristics of each disease. These characters will allow for proper characterization of the teeth that will be used in this study.

Dental caries are known to cause internalized discoloration among a tooth's structure, which can be internalized by the enamel or by the dentin. This discoloration can come from a number of reasons.

One aspect of carry discoloration starts as an initial lesion seen as a white spot, within the tooth. The color differs due the increased porosity of the structure and its refractive index. The hard arrested legion becomes black due to exogenous stain sources. The proteolysis-chelation theory of cavity formation has defined this carious process centered around amino-acid release during proteolysis [4].

Other aspects of discoloration come from the sugar-protein reaction, Maillard reaction (or non-enzymatic browning) and melanin. The Maillard reaction is believed to have products and intermediate chemicals, which push for this theory. In addition, Melanin is a possible candidate, however its location has been scattered amongst the sites of carious teeth deeming it as poor evidence. These two methods have been difficult to distinguish due to their pigments similarity found in spectroscopy [4].

An alternative to these theories would be exogenous co-staining factors. These factors could be dietary chromogens that are entering into the tooth during the carious process amongst microbial communities [4]. This point will be analyzed and assessed further through this study.

C. Characterization of Teeth

The 1000 teeth used for this study were collected and prepared by Colgate. The group received the full set of teeth from three different locations. Those locations include Rutgers School of Dental Medicine, Oral Surgery groups in the New Jersey area, and Therametrics. The teeth used in this study were all taken all due to orthodontic procedures.

All of the teeth were collected and sorted and focused on samples that were of molar teeth. Molar teeth were used to guarantee that there were high enough amounts from each sample to be process through LC-MS/MS and 16s Sequencing. Each tooth was cut to solely sample the crown of each tooth. The root wouldn't provide measurable data that would be of interest of the study.

The teeth were labelled and grouped into three different categories. Those categories include: control, normal teeth, and stained teeth. Control teeth were used as a whiteness standard for each tooth. Normal Teeth were used as the main sample group for this study. There were no treatments done to these teeth to keep them as close to natural as possible. Stained teeth were categorized by the known staining solutions of wine, coffee, tea, and tobacco. These samples were indexed to identify if any of the molecules can be traced to normal natural teeth. The staining procedures are outlined below:

Wine stain: 25 human teeth were, after imaging, placed into a staining apparatus with a pure wine stain broth. The staining apparatus and broth was placed in a 37 C incubator and the teeth cycled in and out of the wine stain broth repeatedly for ~ 2 weeks. The

wine stain broth was refreshed every 48 hours. After the staining period, the teeth were removed, rinsed well, and placed back into 0.1% thymol and refrigerated until they were shipped for analysis.

Coffee stain: 25 human teeth were, after imaging, placed into a staining apparatus with 10% by weight coffee stain broth. The staining apparatus and broth was placed in a 37 C incubator and the teeth cycled in and out of the coffee stain broth repeatedly for ~ 3 weeks. The coffee stain broth was refreshed every 48 hours. After the staining period, the teeth were removed, rinsed well, and placed back into 0.1% thymol and refrigerated until they were shipped for analysis.

Tea stain: 25 human teeth were, after imaging, placed into a staining apparatus with 10% by weight tea stain broth. The staining apparatus and broth was placed in a 37 C incubator and the teeth cycled in and out of the tea stain broth repeatedly for ~ 3 weeks. The tea stain broth was refreshed every 48 hours. After the staining period, the teeth were removed, rinsed well, and placed back into 0.1% thymol and refrigerated until they were shipped for analysis.

Tobacco stain: 25 human teeth were, after imaging, placed into a staining apparatus with 10% by weight tobacco stain broth. The staining apparatus and broth was placed in a 37 C incubator and the teeth cycled in and out of the tobacco stain broth repeatedly for ~ 3 weeks. The tobacco stain broth was refreshed every 48 hours. After the staining

period, the teeth were removed, rinsed well, and placed back into 0.1% thymol and refrigerated until they were shipped for analysis.

Each tooth within the dataset was characterized by Colgate along 4 staining scales. These scales include the L, A, B, and WIO scales. The L scale measures colors from white to black. The A scale measures colors from red to green. The B scale measures colors from yellow to blue. The WIO scales defines the subjective light and darkness of the tooth from the perceived eye. These scales were used to define both extrinsic (enamel) and intrinsic (dentin) stains.

The CIELAB color system is used for measuring tooth color in dentistry. As described above, the “LAB” represents specific hue scales that measures between two distinctive colors. The reading for each tooth was measured by an Easy Shade Device.

.

D. Metabolomic Tools

The main metabolomic analysis computed was done by the Global Natural Products Social Molecular Networking (GNPS) ecosystem. This metabolomic interface is described below:

> Global Natural Products Social Molecular Networking (GNPS) is a web-based mass spectrometry ecosystem that aims to be an open-access knowledge base for community-wide organization and sharing of raw, processed or identified tandem mass (MS/MS) spectrometry data. GNPS aids in identification and discovery throughout the entire life cycle of data; from initial data acquisition/analysis to post publication [6].

The GNPS interface was the staple of this study and allowed for specific analysis of the Molecular Network, Feature-Based Molecular Network, and Network Annotated Propagation. Each tool used in this study is described below.

> Molecular Networks are visual displays of the chemical space present in tandem mass spectrometry (MS/MS) experiments. This approach groups sets of spectra from related molecules (known as molecular families) even when the spectra themselves are not identified (do not match to any known compounds) [6].

Molecular Networking was used to differentiation specific clusters that are represented by the different staining treatment types (Coffee, Wine, Tea, and Tobacco).

Feature-Based Molecular Networking (FBMN) is a computational method that bridges popular mass spectrometry data processing tools for LC-MS/MS and molecular networking analysis on GNPS [7].

Feature-Based Molecular Networking was used to examine specific metabolites and how they effect each color determining scale.

Network Annotated Propagation (NAP) uses spectral networks to propagate information from spectral library matching in order to improve *in silico* fragmentation candidate structure ranking[8].

NAP was used to predict the structures for specific metabolites that aren't within the GNPS database and assist in assessing microbial community interactions from each metabolite.

Statistical Analysis was implemented to notice significant differences within the dataset, highlight important metabolites, and compare important metabolites against each other. The Statistical analysis that took place include: PCoA, PLS-DA, Random Forrest, and Spearman Correlations.

The master table was used to compute the distances using our in-house analysis platform "clusterApp", which is compatible with the PCoA generation software, EMPeror. Canberra and Bray-Curtis dissimilarity distances were used to express correlations found after data analysis. The data was then normalized using data normalization with Probabilistic Quotient Normalization(30%)[11].

The supervised analysis utilizes the information of a priori delineated group to define the separation between them. Analysis was conducted using the Metaboanalyst platform[10]. Partial Least Squares Discriminant Analysis (PLS-DA) [K] was used to determine which metabolites drive the differences among the experimental categories of interest[10].

For the Random Forrest analysis, tables were uploaded to metaboanlyst.ca and filtered using the interquartile range (IQR). Samples then underwent quantile normalization and were auto-scaled (mean-centered and divided by the standard deviation of each variable). The median was used to separate the classifications for each attribute (Ex: L scale White vs Black).

Spearman coefficient correlations was used as a statistical measure of strength within the monotonic relationship between the different metabolites and each specific scale analysis.

E. Research Objectives

My research objective is to explore the fundamental causes of tooth staining from identifying key chemistry, key microbial communities, and the relationships amongst the two groups in regards to staining.

Chapter II

Sample Method Optimization

A. Tooth Collection and Preparation

The tooth collection and preparation process was all managed by Colgate. The process began with the collection of teeth from Therametric, Oral Surgery groups around New Jersey, and Rutgers School of Dentistry. Once at Colgate the teeth were cut at the crown and packaged or stained for further analysis. Teeth were then measured with the CIELAB color system acquired by an Easy Shade Device.

B. Metabolome Processing

900 Human teeth were washed with cold water and dried with Kimwipes. Teeth were then powderized in a tissue lyser (30 sec. / 30 hz) in stainless steel jars (10mL) with stainless steel bead. Each sample was scraped out of the jars using collecting spoon. The spoon and stainless steel jar w/ ball were cleaned with IPA:H_2O (1:1) after each sample. Tooth powder was transferred to 5ml glass vials and barcoded. Each powdered tooth was weighed on scales and the mass values were found to range from 0.2 to 1.5 g. The teeth were then stored at -80°C for bacterial growth prevention before preparation for metabolomic and sequencing plates. Tooth storage solution was stored in -4°C fridge for further analysis for specific samples.

For extraction, ~250mg of each tooth powder were plated. A synthetic standard of hydroxyapatite and control teeth were included in each 96-well deep plate. 49 teeth with a mass below 350mg were excluded from further genomic sequencing and plated solely for metabolomic analysis. 88 tooth samples were plated with two hydroxyapatite controls and two tooth controls for both analyses in separate Eppendorf 2mL 96-well

deep plates. Sample were plated in the biohood after cleaning all the surfaces with isopropanol and an RNase solution. Samples plated for microbial sequencing were stored in -80°C for further analysis.

For metabolomics the teeth were then extracted with 900uL of MeOH:H_2O:EtOAc (1:1:1) with internal standard 1, sulfadimethoxine, using a multichannel pipette, vortexed (~10 sec.), sonicated for 90 minutes and extracted overnight at 4°C. Sample were then centrifuged (2000x rpm) for 20 minutes and transferred 600 uL to a 2mL 96-well deep plate to undergo centrifugal evaporation in lyophilizer. Samples were then resuspended using 100ul MeOH:H_2O (1:1) with internal standard 2, sulfachloropyridizine, vortexed for 10 seconds, sonicated for 15 minutes, and centrifuged (2000x rpm) for 20 minutes. An 80uL aliquot of each sample was then transferred to ThermoScientific 96-well shallow plate with external mix of standards and subjected to LC-MS/MS analysis.

C. Experimental Protocol

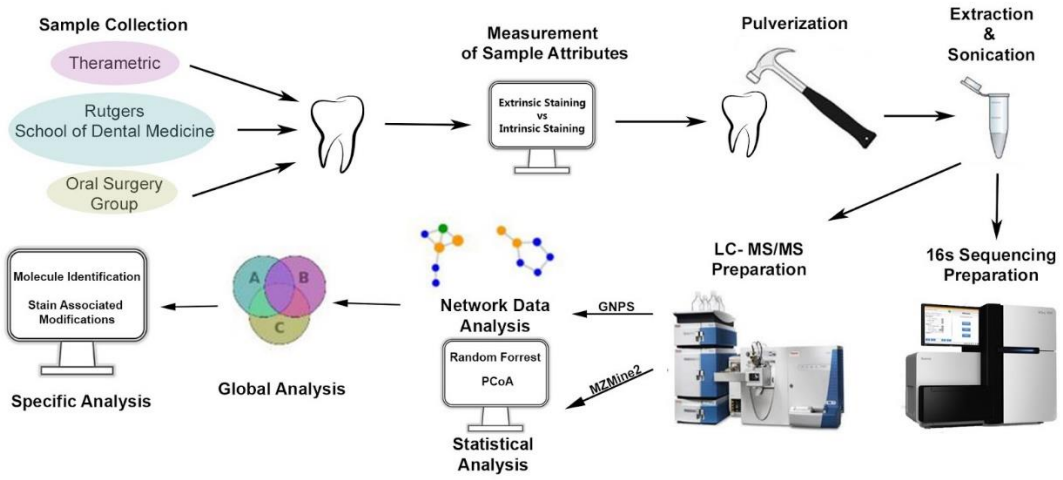


Figure 1. Experiment Protocol
Samples were collected from one of three groups, Therametric, Rutgers School of Dental Medicine, and an Oral Surgery Group. They were then measured and archived based on their color. Both the extrinsic staining and intrinsic staining were measured and sent to UCSD to begin the data collection process. The teeth were pulverized using tissue lyser (30 sec. / 30 hz) in stainless steel jars (10mL) with a stainless steel bead. Extraction, sonication, and lyophilization, took place to prep for both LC-MS/MS and 16s Sequencing in 96-well plates.

D. Instrumentation

Reverse-phase liquid chromatography – mass spectrometry was performed using a Thermo Vanquish UHPLC system coupled to a QExactive Orbitrap mass spectrometer. Data were acquired using data dependent acquisition (*m/z* 80-1200), subsequently fragmenting the five most abundant precursor ions. The teeth were

profiled with Prosolia Flowprobe coupled to LTQ Orbitrap XL with ~500mL of 50% MeOH as extraction solvent used.

The injected samples were chromatographically separated using an Vanquish uHPLC (Thermo Fisher Scientific, Waltham, MA) controlled by Thermo SII for Xcalibur software (Thermo Fisher Scientific, Waltham, MA), using a 100 x 2.1 mm Kinetex 1.7 µM, C18, 100Å chromatography column (Phenomenex, Torrance, CA), 40°C column temperature, 0.5 mL/min flow rate, mobile phase A 99.9% water (J.T.Baker, LC-MS grade) 0.1% formic acid (Thermo Fisher Scientific, Optima LC/MS), mobile phase B 99.9% acetonitrile (J.T.Baker, LC-MS grade) 0.1% formic acid (Fisher Scientific, Optima LC/MS), with the following gradient: 0-1 min 5% B, 1-8 min 100% B, 8-10.9 min 100% B, 10.9-11 min 5% A, 11-12 min 5% B for all samples and benchmarking solutions except for benchmark 1 where the gradient was as follows: 0-1 min 5% B, 1-4.5 min 100% B, 4.5-5.5 min 100% B, 5.5-5.6 min 5% A, 5.6-6 min 5% B.

MS analysis was performed on an Orbitrap (QExactive, Thermo Fisher Scientific, Waltham, MA) mass spectrometer equipped with HESI-II probe sources and controlled by Xcalibur 3.0 software. The following probe settings were used for both MS for flow aspiration and ionization: Spray voltage of 3500 V, Sheath gas (N2) pressure of 35 psi, Auxiliary gas pressure (N2) of 10 psi, ion source temperature of 270°C, S-lens RF level of 50 Hz and Aux gas heater temp. at 440°C.

For orbitrap MS, spectra were acquired in positive ion mode over a mass range of 100-1500 m/z. An external calibration with Pierce LTQ Velos ESI positive ion

calibration solution (Thermo Fisher Scientific, Waltham, MA) was performed prior to data acquisition with error rate less than 1 ppm. Data acquisition parameters were set as follows: minutes 0-0.5 were sent to waste; minutes 0.1-12 were recorded with data-dependent MS/MS acquisition mode. Full scan at MS1 level was performed with resolution of 35K in profile mode. The 10 most intense ions with 2 m/z isolation window with m/z 0.5 offset per MS1 scan were selected and subjected to normalized collision induced dissociation with 30eV. MS2 scans were performed at 17.5K resolution with max IT time of 60ms in profile mode. MS/MS active exclusion parameter was set to 5.0 s.

Data Processing

The LC/MS raw data files were converted to mzXML format and feature detection was performed with the MZmine2 software[12.] The software settings were as follows. Mass detection was performed with a signal threshold of 1.0E3 for MS1 and 1.0E2 for MS2. For the chromatogram building, the mass tolerance was set to 10 ppm, the minimum peak time span .01s, and minimum height 5.0E3. For chromatographic deconvolution, the local minimum search algorithm was used; m/z range for MS2 scan pairing was set at 0.025Da and RT at 0.1 min. range. The peaks were de-isotoped within 25 ppm m/z and 0.2 min RT tolerances, aligned, gap-filled using the same tolerances and then filtered to retain only peaks that appear in at least 2 samples with minimum 2 peaks in isotope pattern to create the feature table. Peaks present in any of the blanks

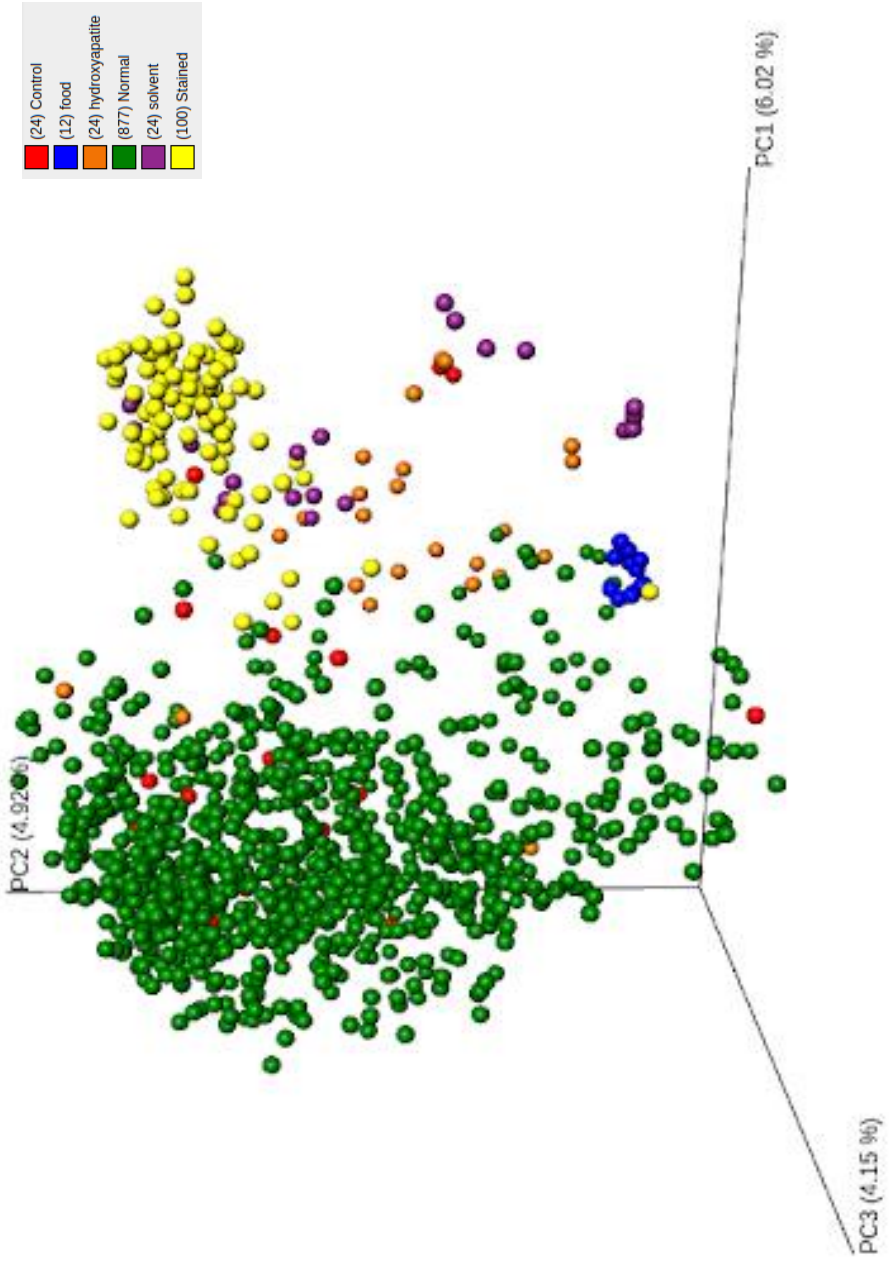
were removed from the final feature table unless at least one sample contained the peak at abundance 3x or above.

The Molecular features (ions) were extracted using mzMine2 software. MzMine2 is an open source software for mass spectrometry with data analysis. The MS1-based feature detection was performed using parameters suitable for MS platform used (Q-Exactive) with the parameters listed in the paragraph above. The Aligned feature table was then exported with the most important metadata categories for inner and outer staining, these include the L, A, B, and WIO whiteness measurement system. Then they were combined in RStudio(R) with the feature table, which was used to create a master table for further statistical analysis. After several feature tables with single important metadata categories were exported for further correlation analysis using metaboanalyst.ca.

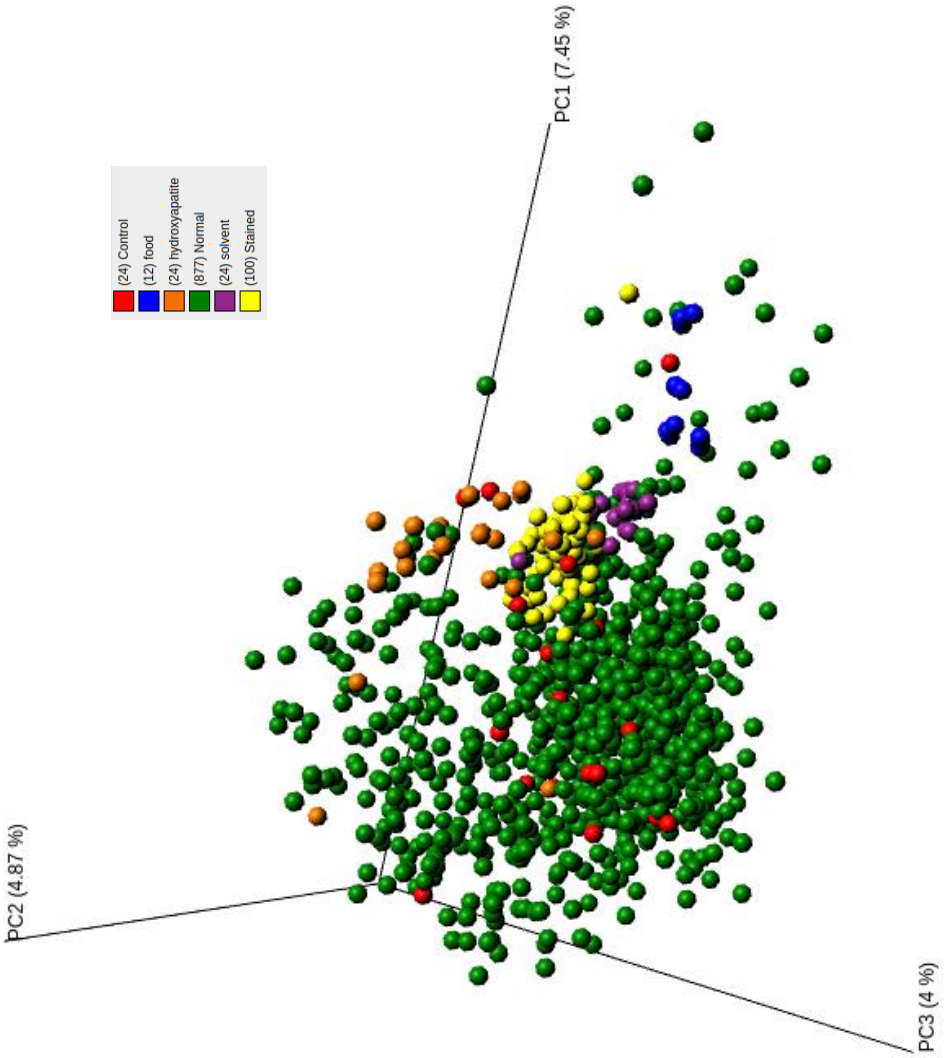
Chapter III

Results

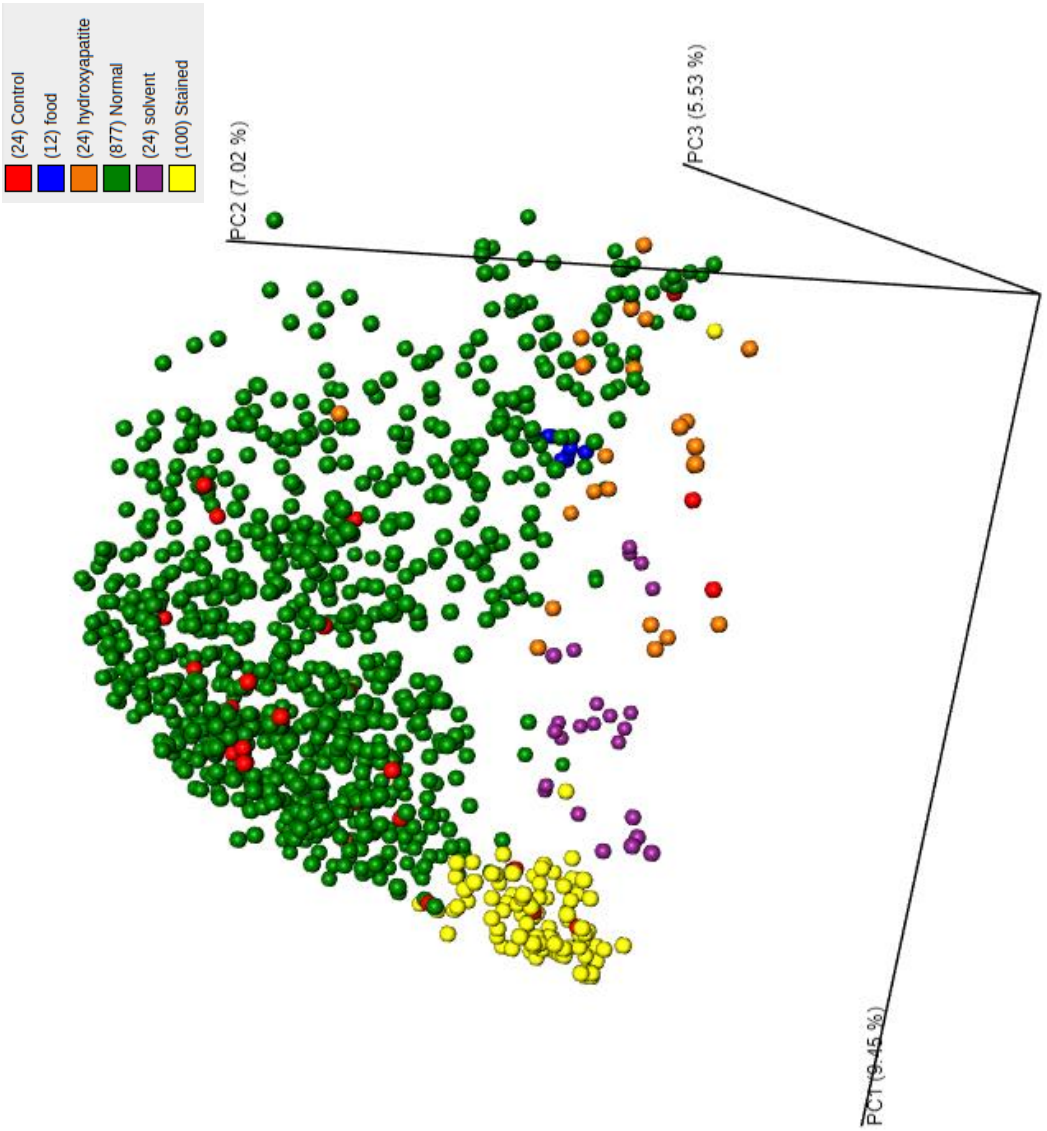
Figure 2. MS1 based PCoA, Distance Matrices. PQN normalization.
Untargeted data analysis on MS/MS data from teeth extracts, hydroxyapatite controls, and staining solutions. a) Principal coordinate analysis plot (PCoA) constructed from MS/MS data and spectral counts as a proxy for signal intensity with binary Jaccard distance metric. b) and c) PCoA of all samples computed with Canberra and Bray-Curtis distance metrics respectively.


(24) Control
(12) food
(24) hydroxyapatite
(877) Normal
(24) solvent
(100) Stained
PC1 (6.02 %)
PC2 (4.92 %)
PC3 (4.15 %)

a)


(24) Control
(12) food
(24) hydroxyapatite
(877) Normal
(24) solvent
(100) Stained
PC1 (7.45 %)
PC2 (4.87 %)
PC3 (4 %)

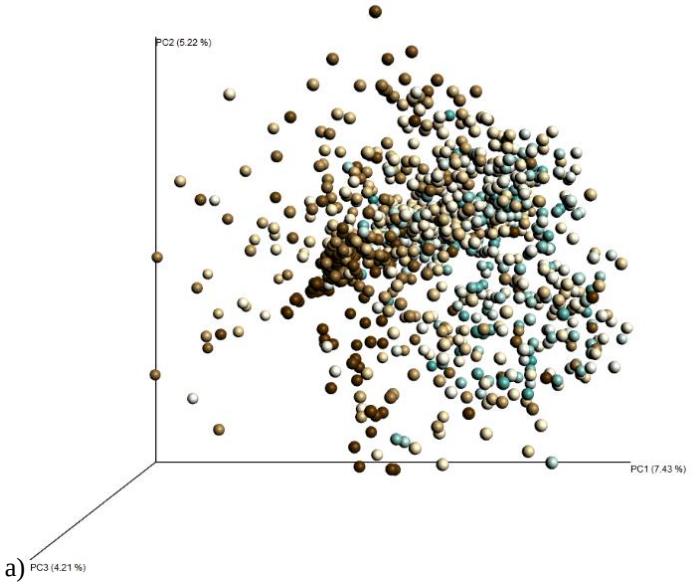
b)


(24) Control
(12) food
(24) hydroxyapatite
(877) Normal
(24) solvent
(100) Stained
PC2 (7.02 %)
PC3 (5.53 %)
PC1 (9.45 %)

c)

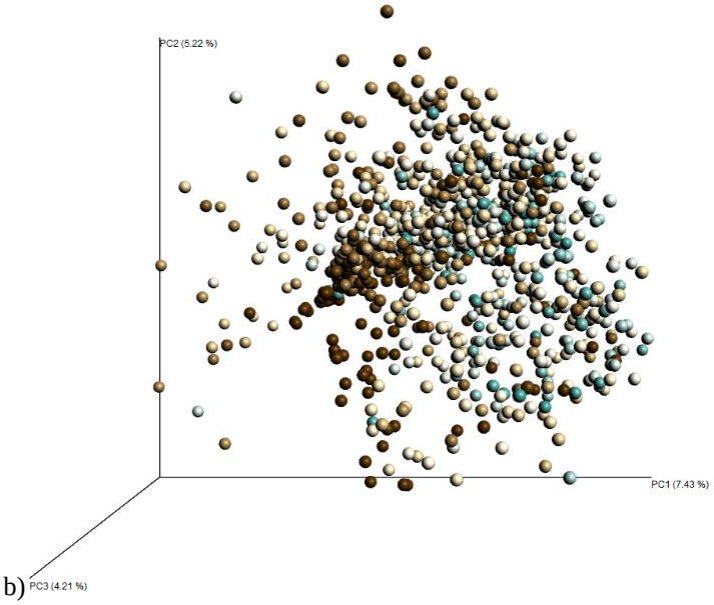
The legend from *Figure 2* is described below:

The Control group (Red) refers to the group of teeth provided by Colgate as a preliminary basis to show that the given samples represent a broad basis. The Staining Solutions group (Blue) refers to the actual staining solutions and agents that were introduced to the teeth which include three replicates of each wine, tobacco, coffee, and tea. The Hydroxyapatite group (Orange) contained samples of the compound hydroxyapatite in powder form, which serves as a negative control against normal teeth due to its enamel like nature. The Normal group (Green) refers to teeth removed from individuals without any prior knowledge or bias on the consumption of any of the staining agents. The solvent group (purple) refers to replicate samples of solvent used for resuspension/extraction later used for mass spectrometry analysis. The Stained teeth group (Yellow) refers to 100 teeth that were artificially stained with the four staining agents that are represented as food.



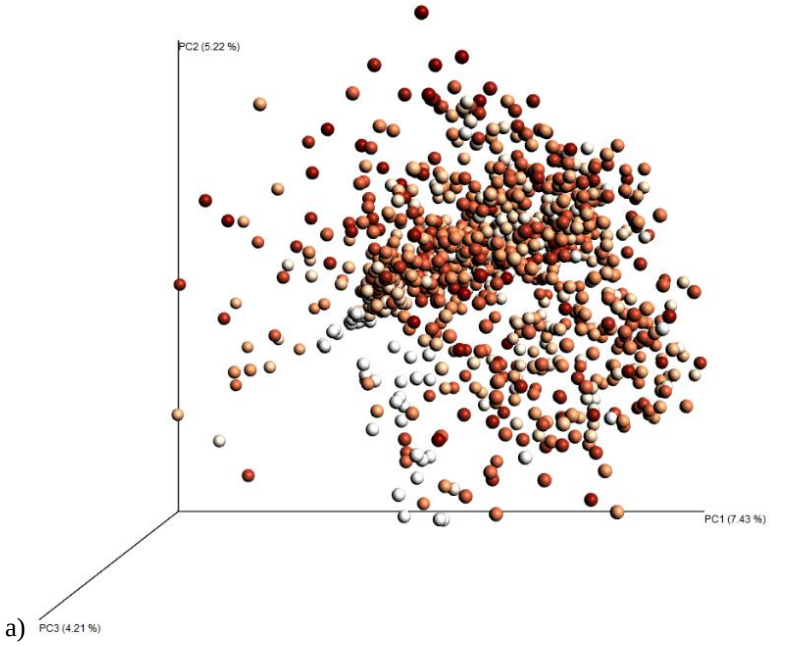
a)

.

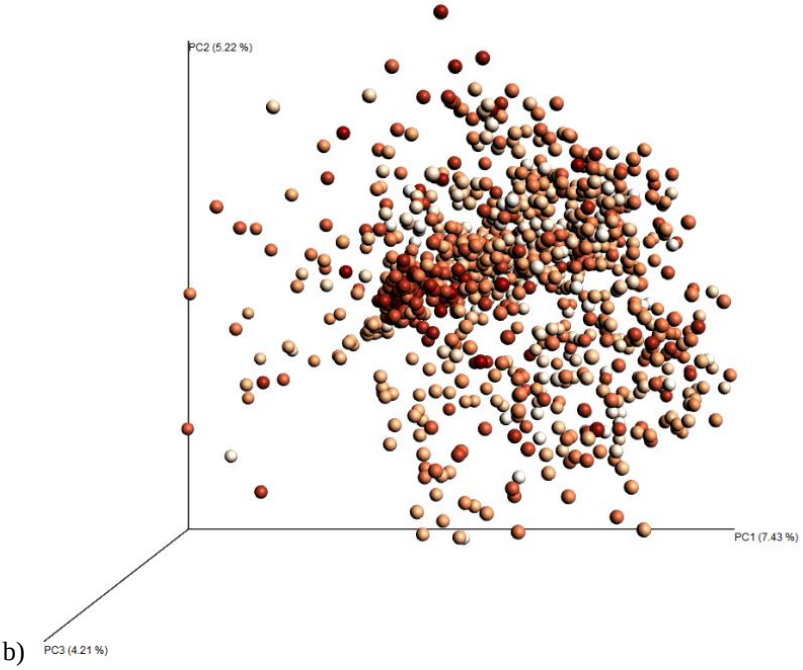


b)

Figure 3. MS1 based PCoA Canberra with PQN normalization of L Scale Untargeted data analysis on MS/MS data from teeth extracts. PCoA of all samples computed with Canberra distance and color-coded based on L scale. The light blue color indicates a higher L and whiter teeth. Brown represents a lower L value and black teeth a) Extrinsic (enamel) and b) Intrinsic (dentin) both show a gradient from PC1 to PC3.



a)



b)

Figure 4. MS1 based PCoA Canberra with PQN normalization of A Scale Untargeted data analysis on MS/MS data from teeth extracts. PCoA of all samples computed with Canberra distance and color-coded based on A scale. The red color indicates a higher A and a more red hue on the teeth. Orange represents a lower A value and a green hue on the teeth a) Extrinsic (enamel) and b) Intrinsic (dentin) both show a gradient of PC3 to PC1.

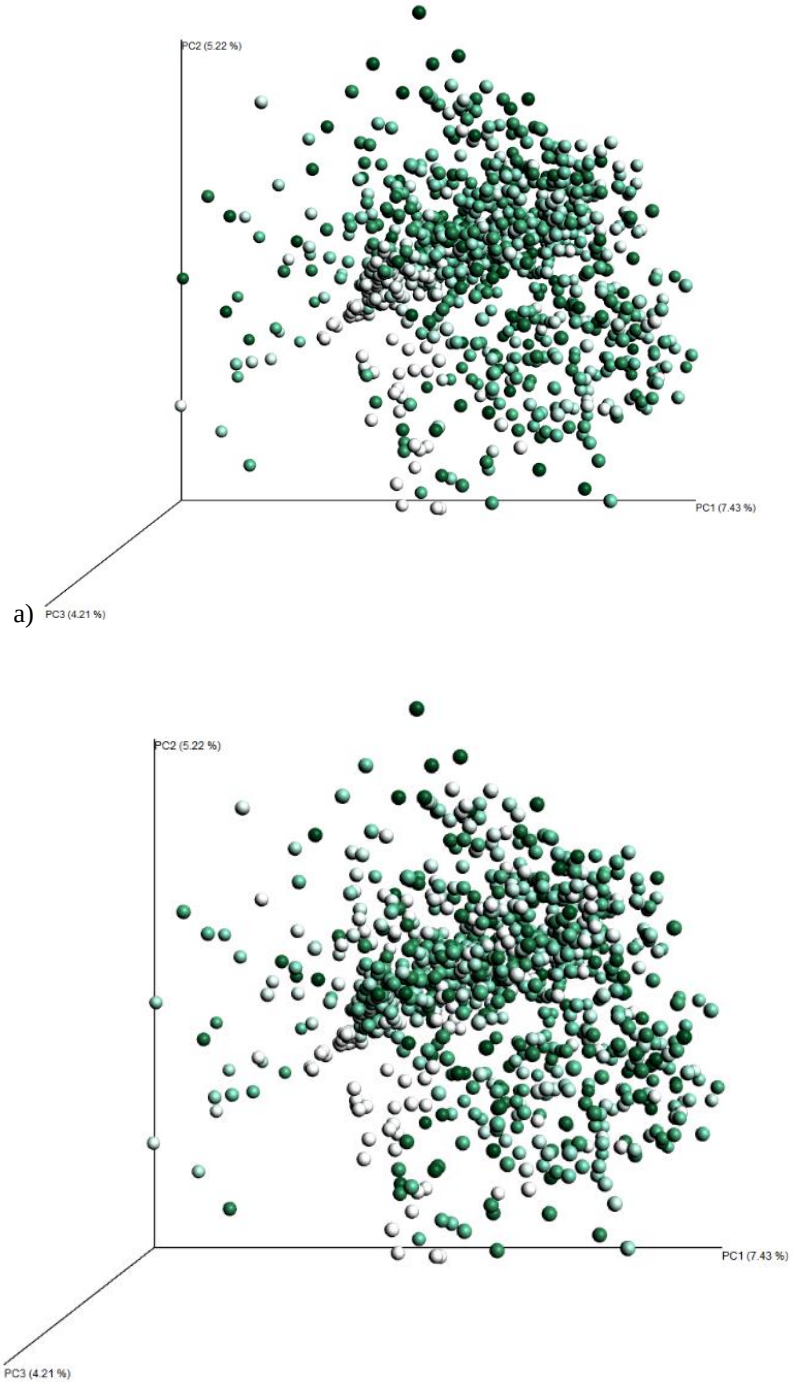


Figure 5. MS1 based PCoA Canberra with PQN normalization of B Scale Untargeted data analysis on MS/MS data from teeth extracts. PCoA of all samples computed with Canberra distance and color-coded based on the B scale. The color green indicates a higher B and a more yellow hue on the teeth. Light blue represents a lower B value and a blue hue on the teeth a) Extrinsic (enamel) and b) Intrinsic (dentin) both show a gradient for PC1 to PC2.

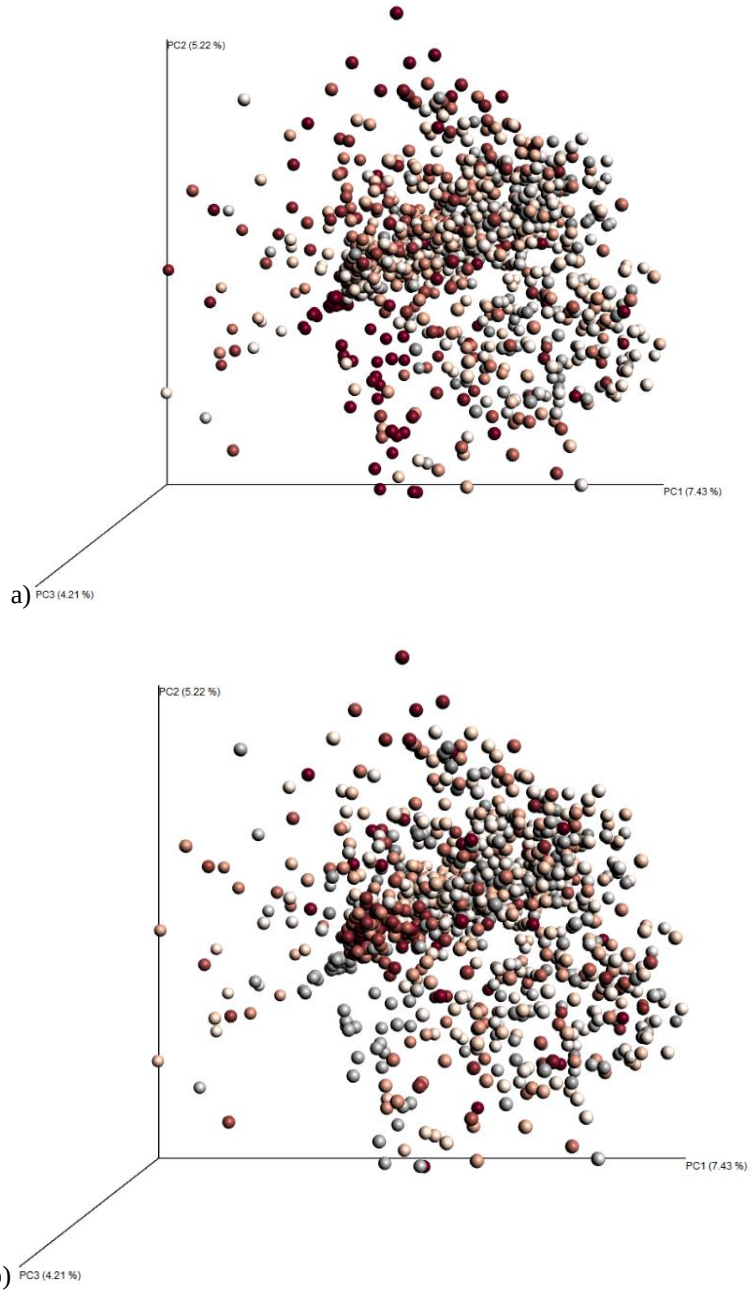
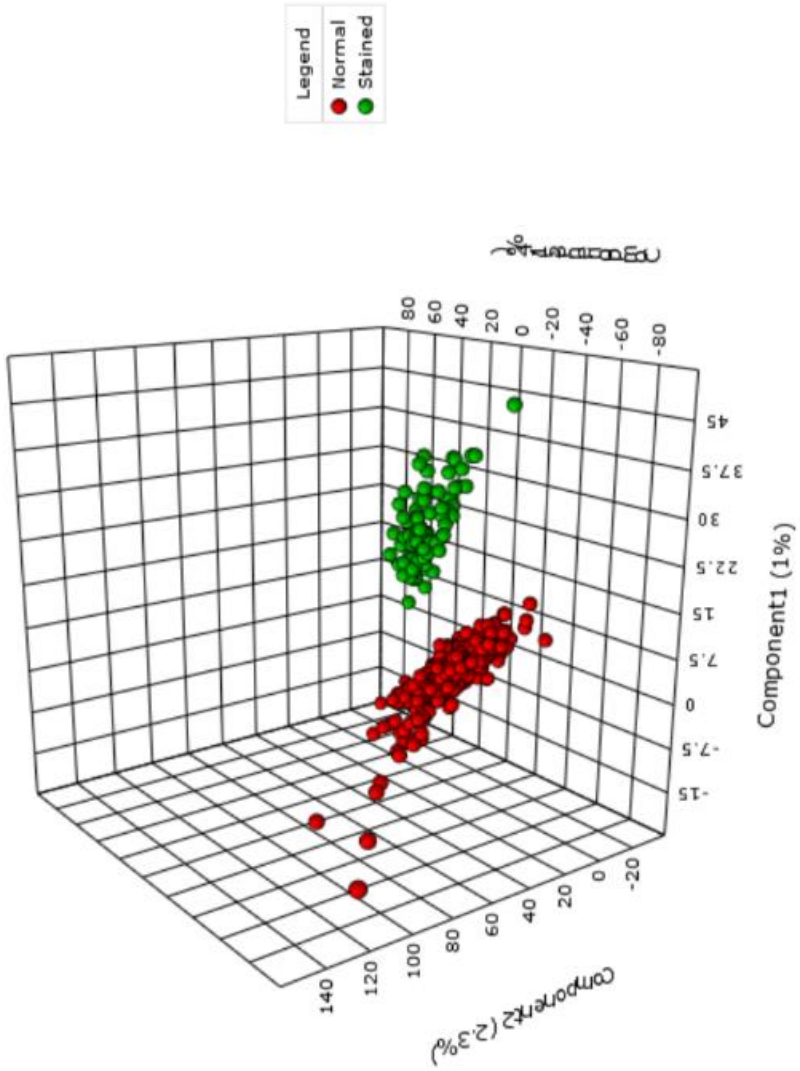


Figure 6. MS1 based PCoA Canberra with PQN normalization of WIO Scale Untargeted data analysis on MS/MS data from teeth extracts. PCoA of all samples computed with Canberra distance and color-coded based on the WIO scale. The grey color indicates a higher WIO and a more white perception of the teeth. Red represents a lower WIO value and a darker color tooth is perceived a) Extrinsic (enamel) and b) Intrinsic (dentin) both show a gradient from PC3 to PC1.

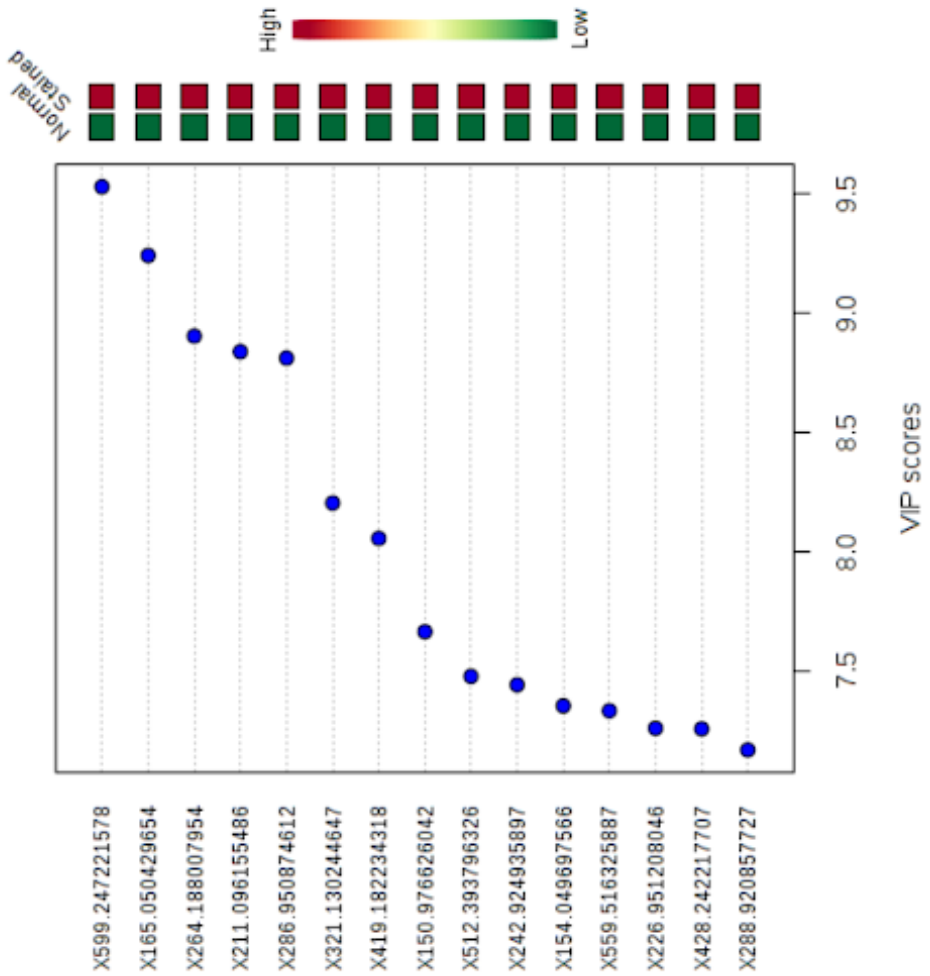
The unsupervised statistical analysis consisted of solely normal teeth. Minute trends can be seen for the A, B, and WIO CIELAB color scales for the analysis, *Figures 4 - 6*. A clear trend for the L CIELAB color scale, *Figure 3*, provides evidence to look into this color matrix more in-depth. However, with little to no discrepancy between the extrinsic and intrinsic stains, there's doubt that there will be major conflicting metabolites between the enamel and the dentin. Further investigation into the intrinsic staining was done using spearman correlation analysis.

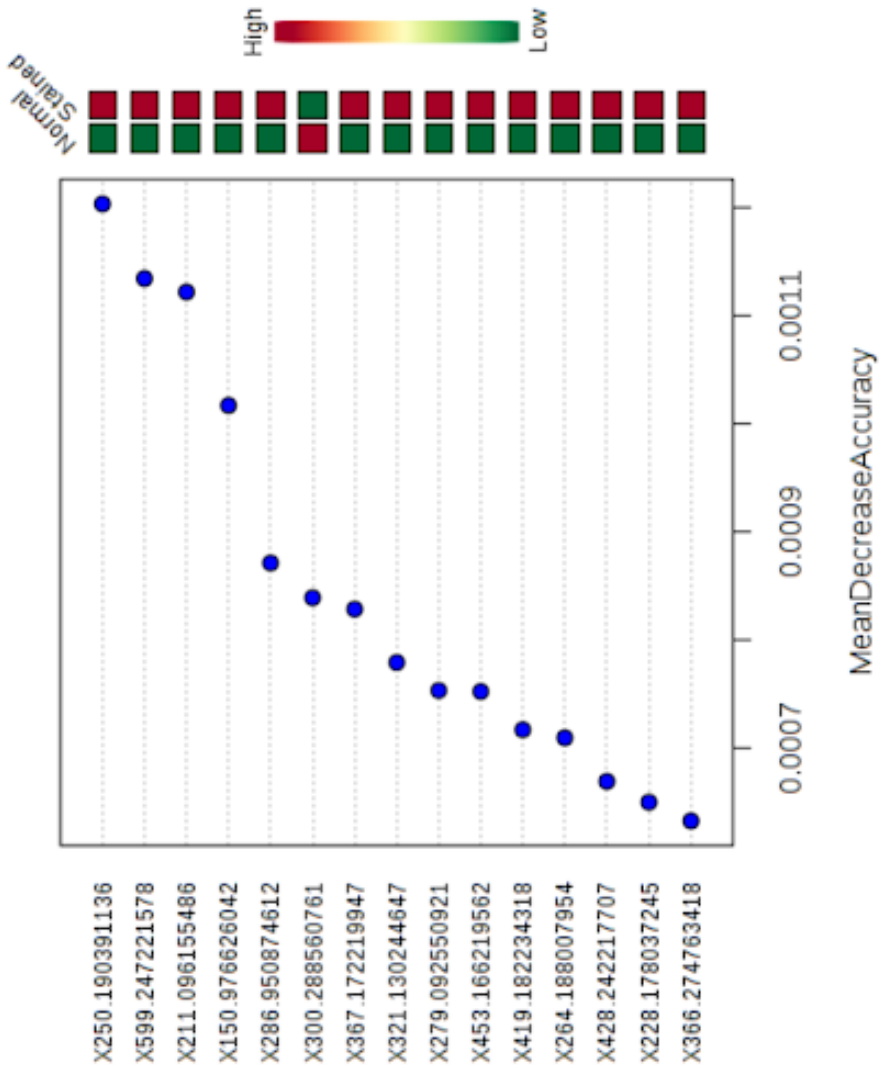
Figure 7. Partial Least Square Discriminant Analysis (PLS-DA) of Normal and Stained teeth. a) 3D plot showing distinct robust separation can be seen between Normal and Stained teeth (Each dot is a sample) b) Top metabolites (features) driving separation between normal and stained teeth c) Random Forest results expressing top contributing metabolites with Mean Decrease Accuracy (MDA)


Legend
Normal
Stained
Component1 (1%)
45
37.5
30
22.5
15
7.5
0
-7.5
-15
80
60
40
20
0
-20
-40
-60
-80
140
120
100
80
60
40
20
0
-20

a)

b)


High
Low
Normal
Stained
9.5
9.0
8.5
8.0
7.5
VIP scores
X599.247221578
X165.050429654
X264.188007954
X211.096155486
X286.950874612
X321.130244647
X419.182234318
X150.976626042
X512.393796326
X242.924935897
X154.049697566
X559.516325887
X226.951208046
X428.242217707
X288.920857727


High
Low
Normal
Trained
X250.190391136
X599.247221578
X211.096155486
X150.976626042
X286.950874612
X300.288560761
X367.172219947
X321.130244647
X279.092550921
X453.166219562
X419.182234318
X264.188007954
X428.242217707
X228.178037245
X366.274763418
0.0007
0.0009
0.0011
MeanDecreaseAccuracy

c)

Due to the gradient seen in the L scale, further supervised statistical analysis took place before addressing the rest of the CIELAB color scales. The robust separation between the normal and stained teeth infers that there are specific metabolites driving these differences. The top features driving the difference for stained teeth can be found from the VIP score and MDA scores in *Figure 7*. Metabolites 599.247221578 and 211.096155486 are expressed from both the PLS-DA and Random forest analysis. The features will be explored further with the GNPS database and via molecular networking.

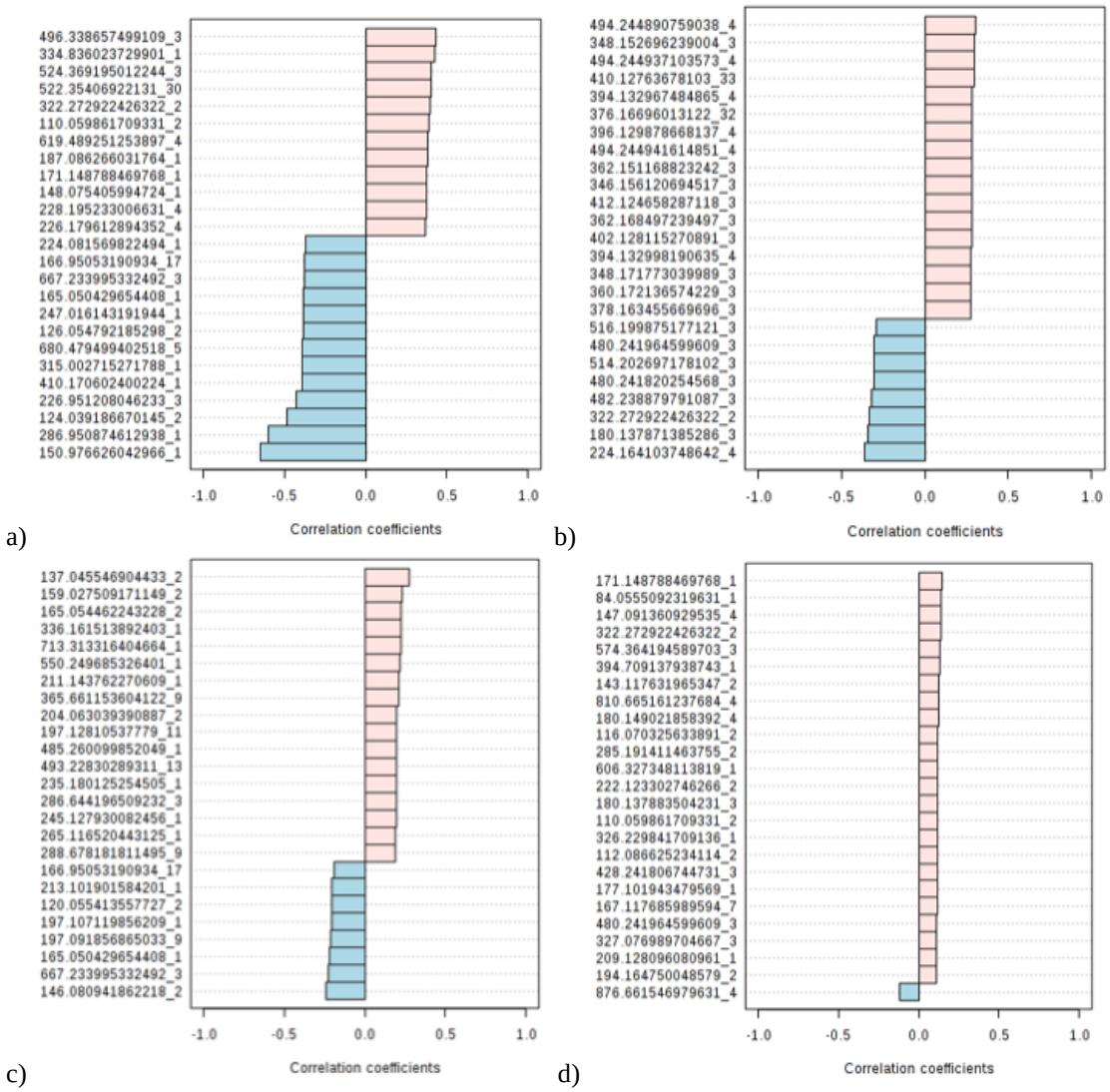


Figure 8. Top Staining Metabolites Contributing to Extrinsic L, A, B, WIO Spearman Correlations.

a) Expresses L scale metabolites 496.338657499109, 334.836023729901, and 524.369195012244 as top metabolites contributing to staining b) Expresses A scale metabolites 494.244890759038, 348.152696239004, and 494.244937103573 as top metabolites contributing to staining c) Expresses B scale metabolites 137.045546904433, 159.027509171149, and 165.054462243228 as top metabolites contributing to staining d) Expresses WIO scale metabolites 171.148788469768, 84.055092319631, and 147.091360929535 as top metabolites contributing to staining

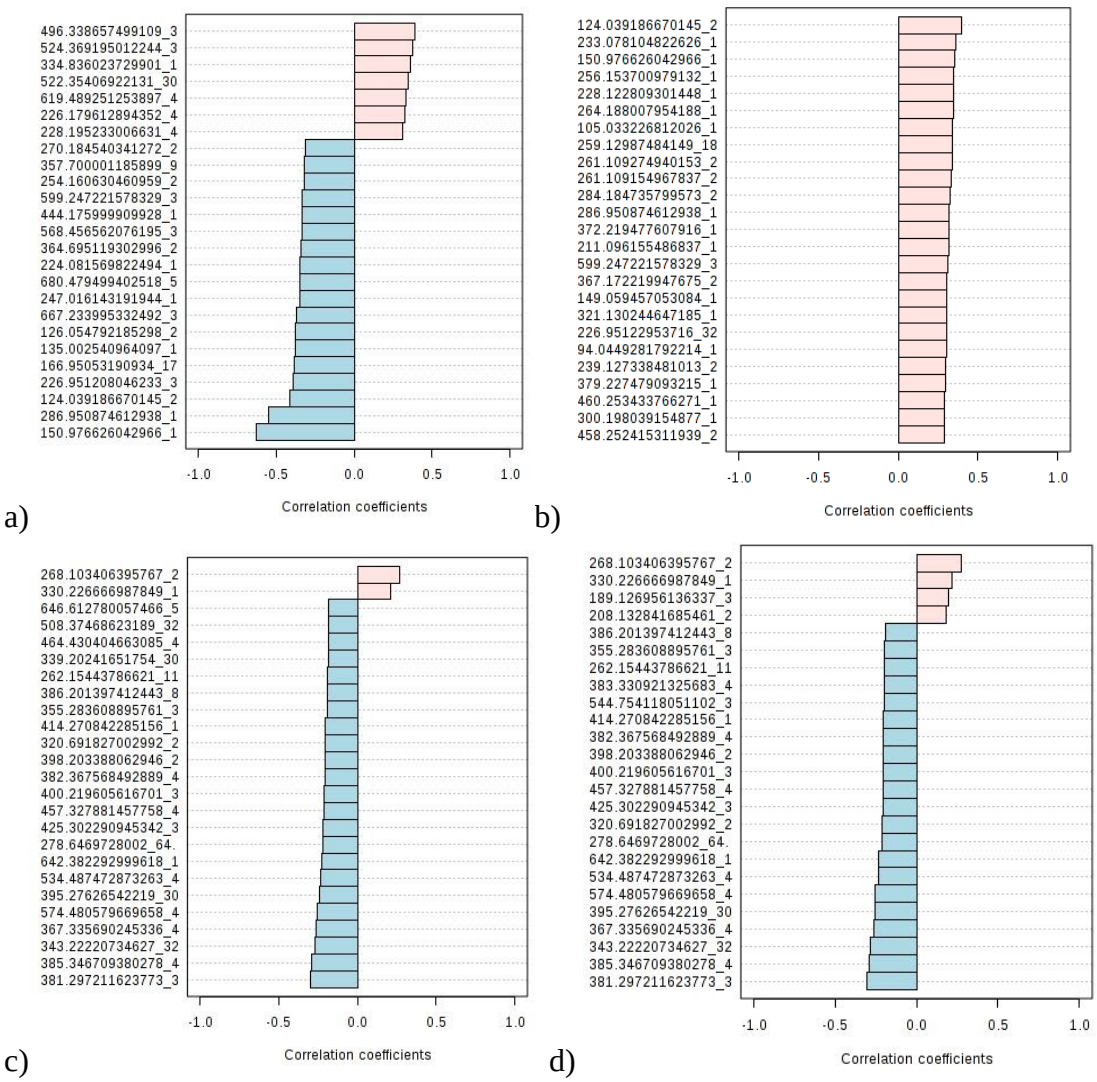


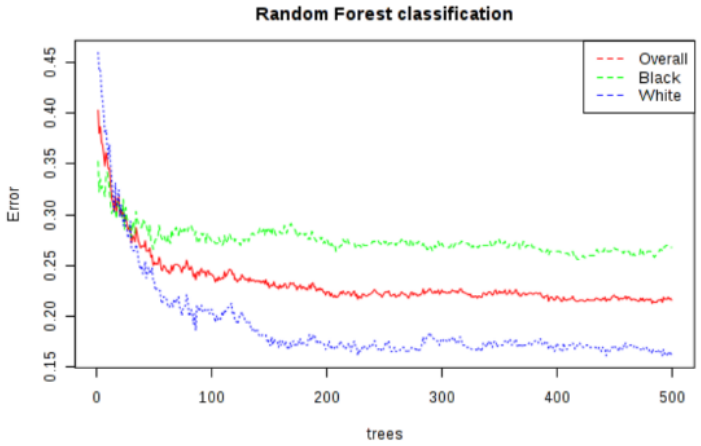
Figure 9. Top Staining Metabolites Contributing Intrinsic L, A, B, WIO Spearman Correlations.

a) Expresses L scale metabolites 496.338657499109, 334.836023729901, and 524.369195012244 as top metabolites contributing to staining b) Expresses A scale metabolites 494.244890759038, 348.152696239004, and 494.244937103573 as top metabolites contributing to staining c) Expresses B scale metabolites 137.045546904433, 159.027509171149, and 165.054462243228 as top metabolites contributing to staining d) Expresses WIO scale metabolites 171.148788469768, 84.055092319631, and 147.091360929535 as top metabolites contributing to staining

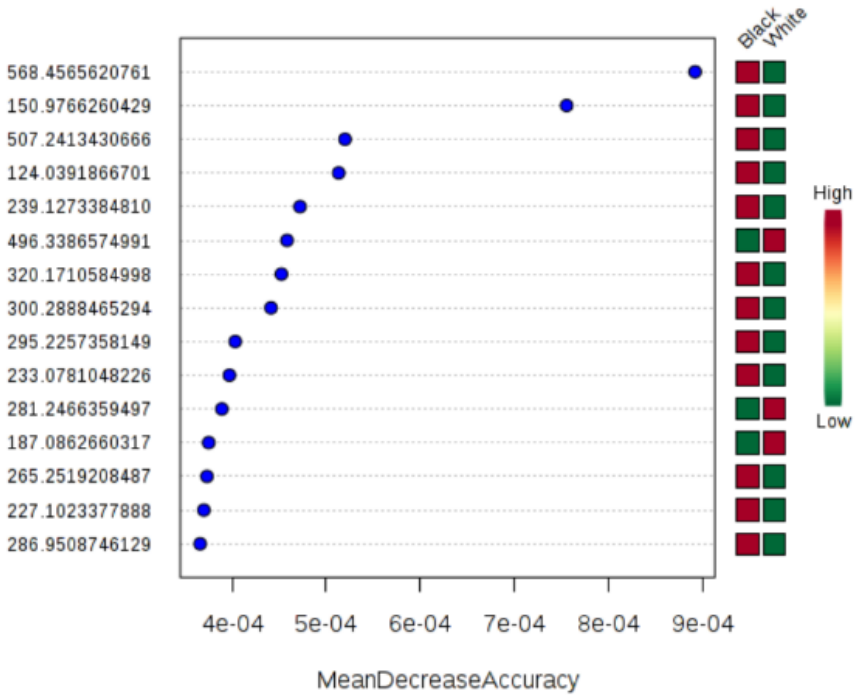
Defining the top contributing metabolites between each color scale allows for clear comparison amongst the different metabolites. Metabolites that affect more than one category are of specific interest to narrow down where this metabolite stems from. Random Forest analysis provides a threshold value to define the characteristics of a given staining category. This analysis gives the highest likelihood of a particular metabolite effecting a specific color stain.

The OOB error is 0.216

	Black	White	class.error
Black	372	136	0.268
White	78	406	0.161



a)

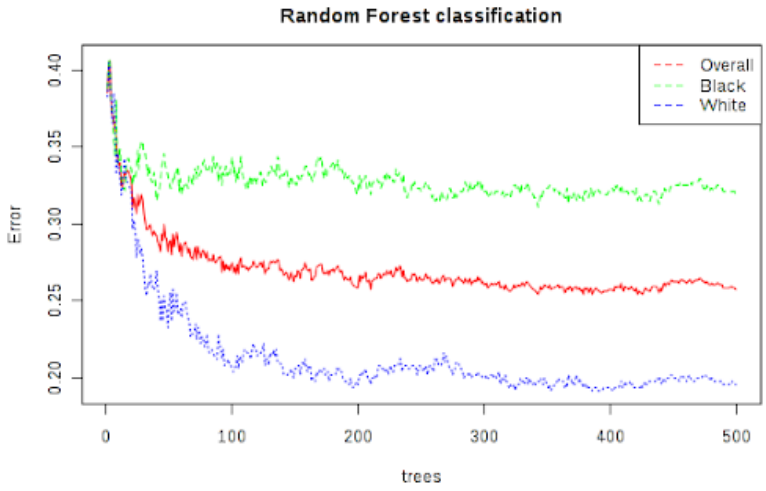


b)

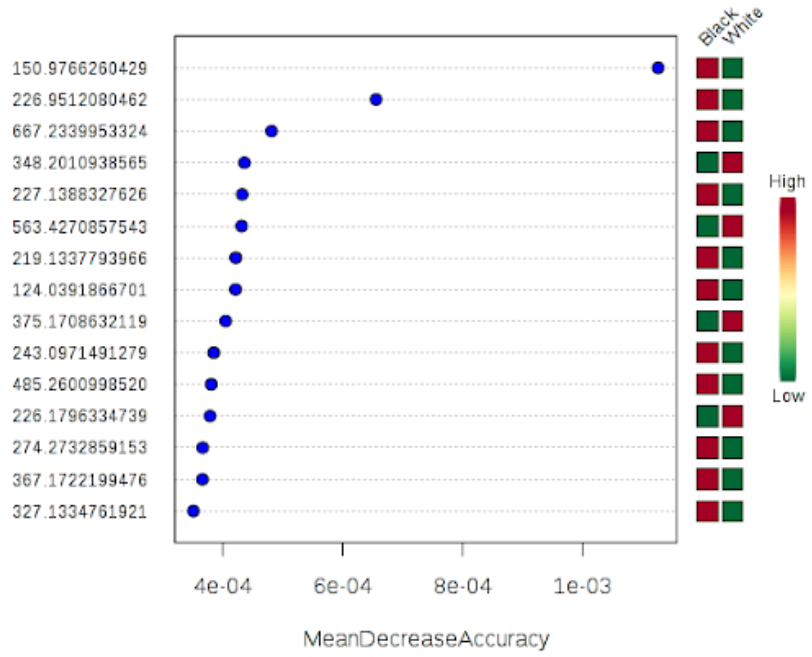
Figure 10. Extrinsic (Enamel) Random Forest Analysis of the L scale
a) For L values, the Black classification has a higher class error than the White classification. Despite the discrepancy between black and white, the class errors are relatively accurate to define these characteristics. b) Metabolites 568.4565620761 and 150.9766260429 had the highest impact on black stains for the L scale

The OOB error is 0.257

	Black	White	class.error
Black	331	155	0.319
White	95	391	0.195



a)

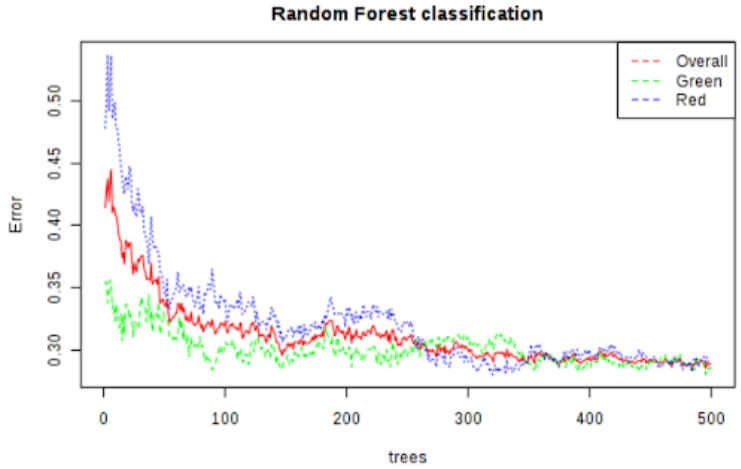


b)

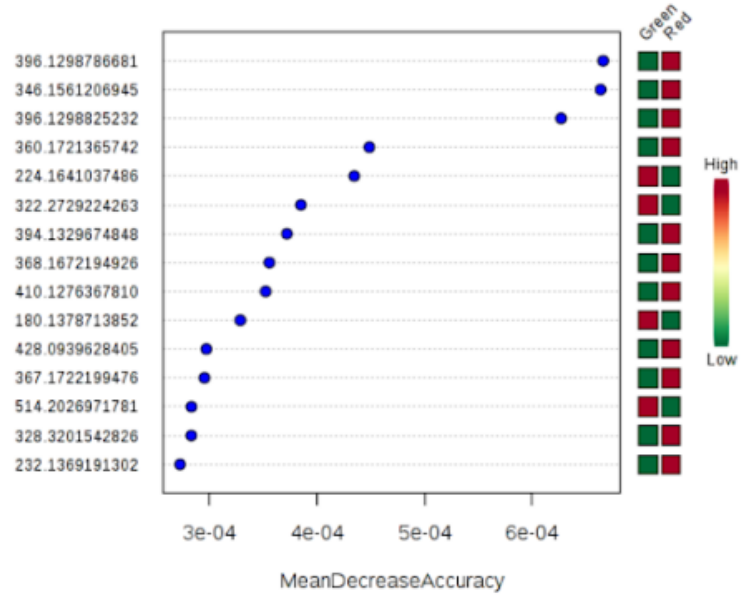
Figure 11. Intrinsic (Dentin) Random Forest Analysis of the L scale
a) For L values, the Black classification has a higher class error than the White classification. Despite the discrepancy between black and white, the class errors are relatively accurate to define these characteristics. b) Metabolites 226.9512080462 and 150.9766260429 had the highest impact on black stains for the L scale

The OOB error is 0.289

	Green	Red	class.error
Green	354	142	0.286
Red	145	351	0.292



a)

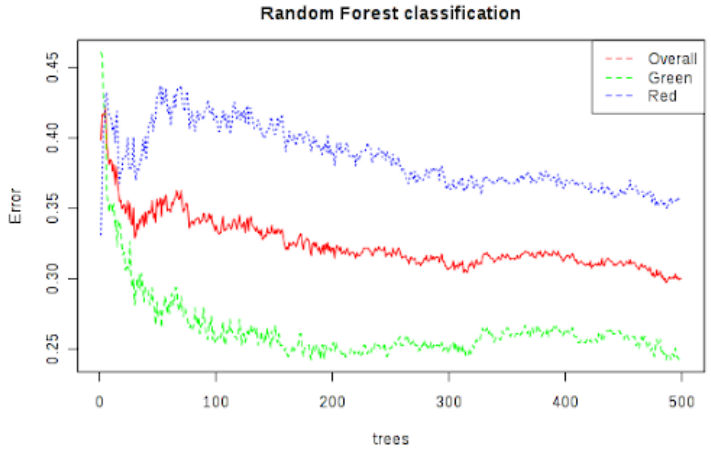


b)

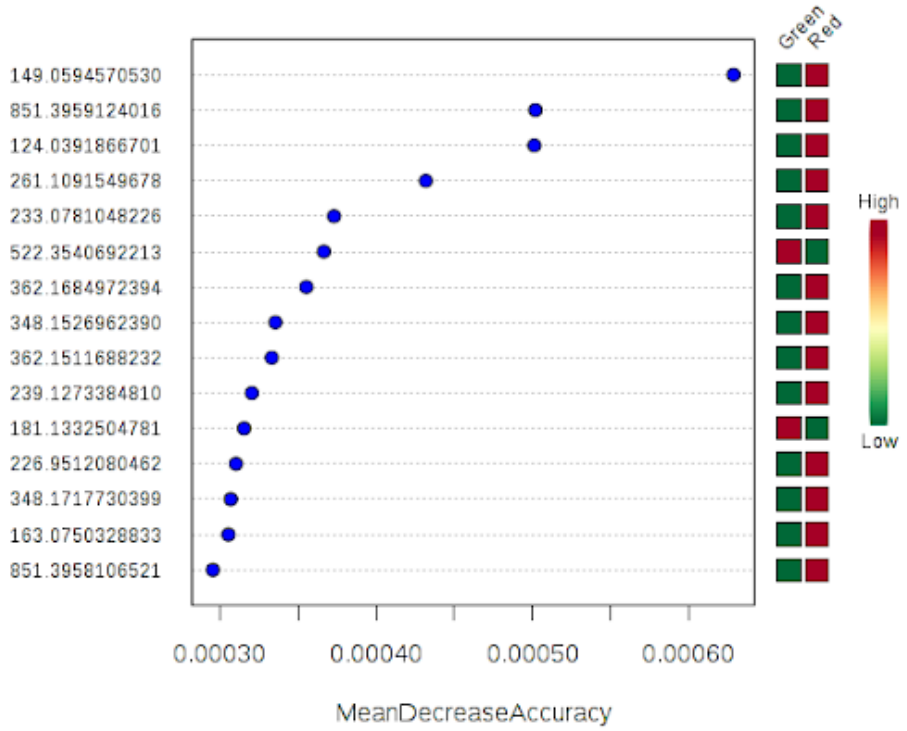
Figure 12. Extrinsic (Enamel) Random Forest Analysis of the A scale
a) For A values, the Red classification has a slightly higher class error than the Green classification. Defining red and green was accurate for the A scale. b) Metabolites 396.1298786681, 346.1561206945, and 396.129882532 had the highest impact on red stains for the A scale

The OOB error is 0.3

	Green	Red	class.error
Green	368	119	0.244
Red	174	315	0.356



a)

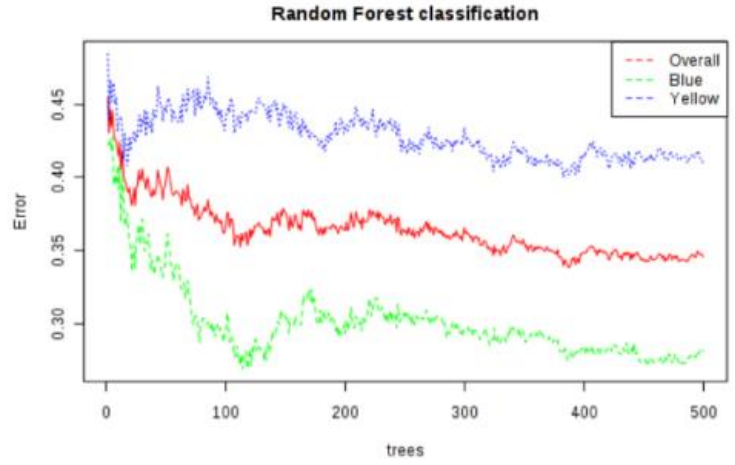


b)

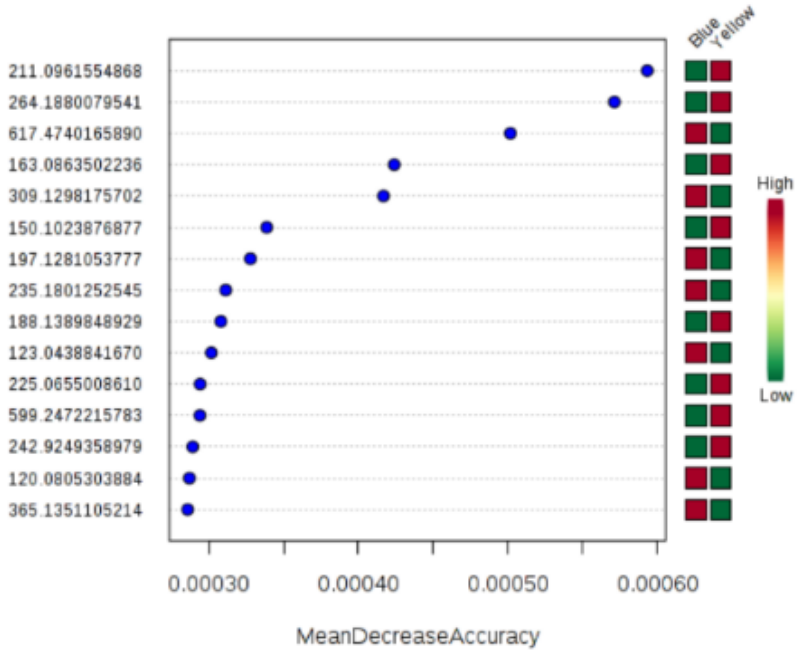
Figure 13. Intrinsic (Dentin) Random Forest Analysis of the A scale
a) For A values, the Red classification has a slightly higher class error than the Green classification. Defining red and green was accurate for the A scale. b) Metabolites 149.0594570530, 851.3959124016, and 124.0391866701 had the highest impact on red stains for the A scale

The OOB error is 0.345

	Blue	Yellow	class.error
Blue	356	139	0.281
Yellow	203	292	0.41



a)

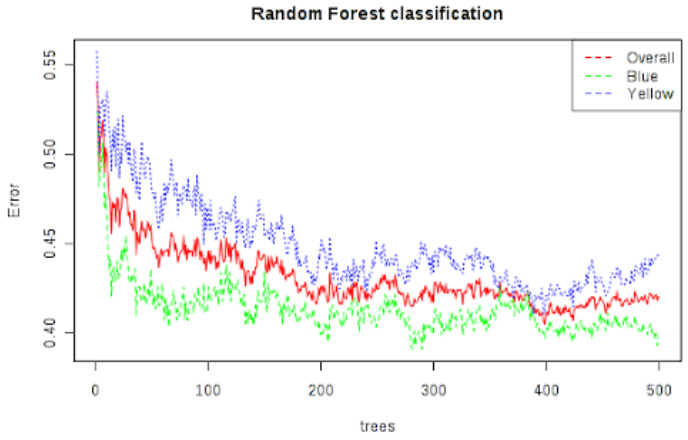


b)

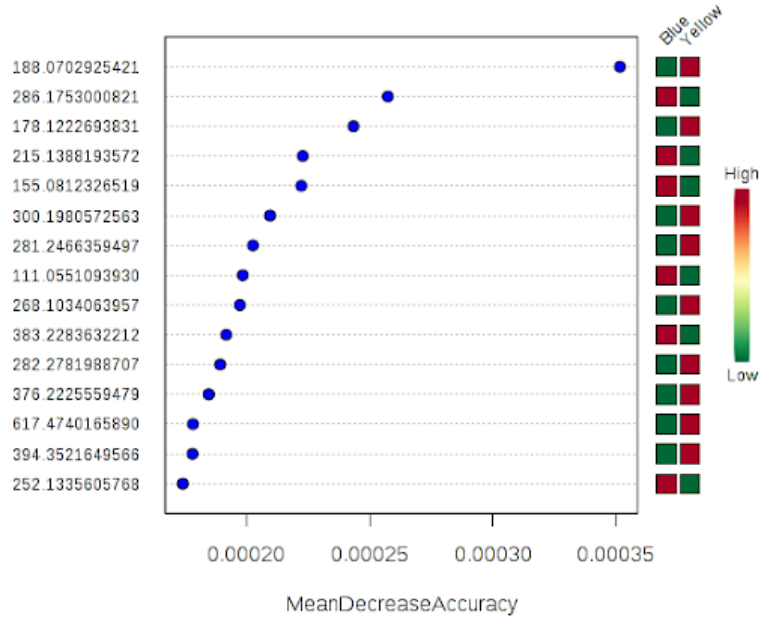
Figure 14. Extrinsic (Enamel) Random Forest Analysis of the B scale
a) For A values, the Yellow classification has a higher class error than the Blue classification. This error less accurately defines the yellow characteristic b) Metabolites 211.0961554868 and 264.1880079541 had the highest impact on yellow stains for the B scale

The OOB error is 0.42

	Blue	Yellow	class.error
Blue	293	193	0.397
Yellow	216	271	0.444



a)

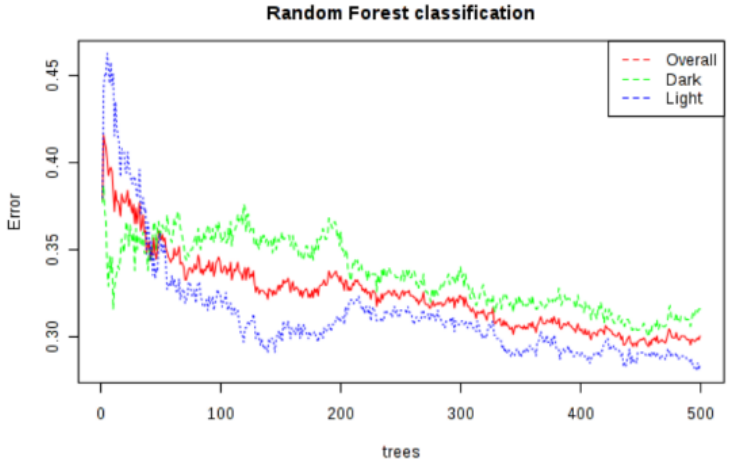


b)

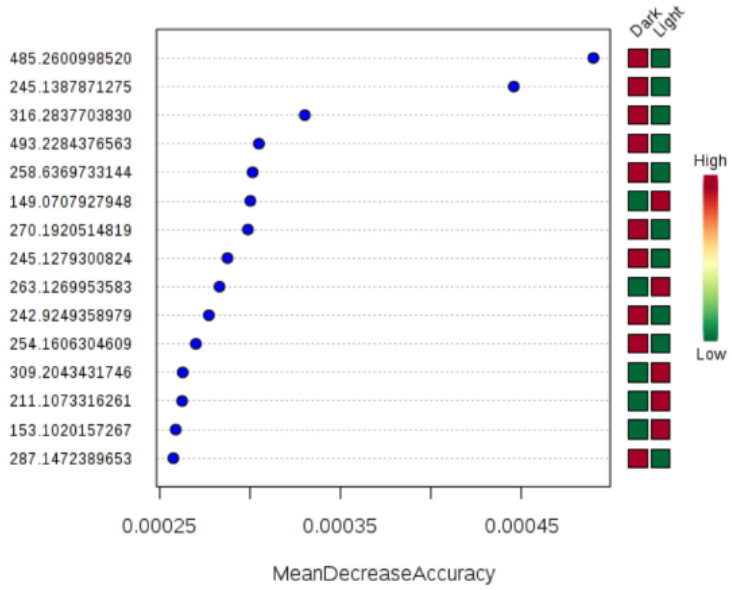
Figure 15. Intrinsic (Dentin) Random Forest Analysis of the B scale
a) For A values, the Yellow classification has a higher class error than the Blue classification. This error less accurately defines the yellow characteristic b) Metabolite 188.0702925421 had the highest impact on yellow stains for the B scale

The OOB error is 0.3

	Dark	Light	class.error
Dark	340	157	0.316
Light	141	354	0.285



a)

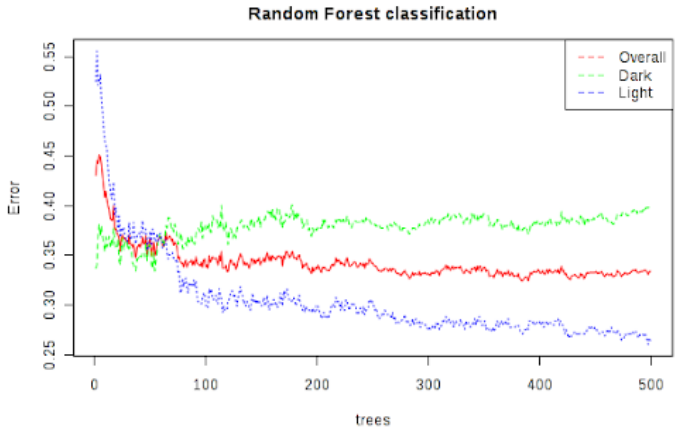


b)

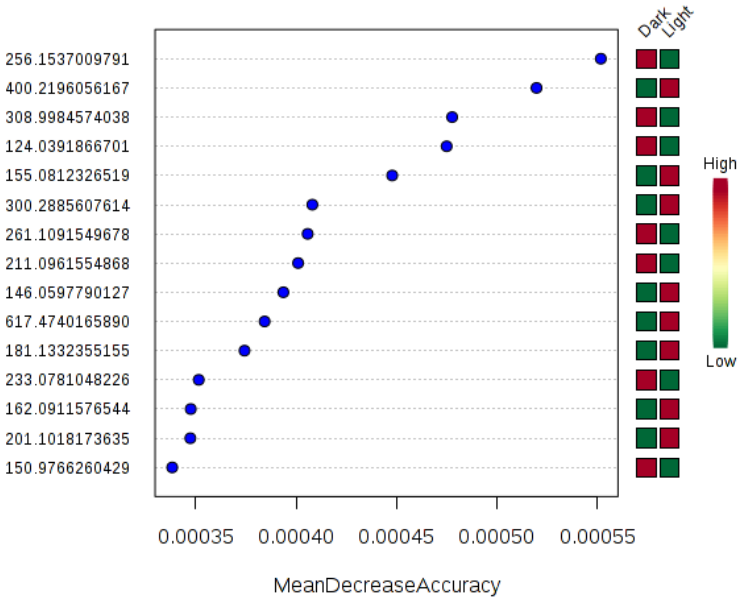
Figure 16. Extrinsic (Enamel) Random Forest Analysis of the WIO scale
a) For A values, the Dark classification has a higher class error than the Light classification. This error less accurately defines the dark characteristic slightly b) Metabolites 485.2600998520 and 245.1387871275 had the highest impact on the perception of darker teeth for the WIO scale.

The OOB error is 0.334

	Dark	Light	class.error
Dark	291	194	0.4
Light	123	342	0.265



a)



b)

Figure 17. Intrinsic (Dentin) Random Forest Analysis of the WIO scale
a) For A values, the Dark classification has a higher class error than the Light classification. This error less accurately defines the dark characteristic b) Metabolites 400.2196056167 and 256.1537009791 had the highest impact on the perception of darker teeth for the WIO scale.

Table 1. Random Forest Variance of Importance Correlations. Top metabolite matches with spearman correlations have been highlighted in a hue of orange. The darker the hue the more importance it has to staining overall.

Extrinsic & Intrinsic Matched Features	Matched Classifications	Variance of Importance	Variance of Importance	Variance of Importance	Variance of Importance
367.172219947675_244.442006085193	A & WIO (INTRINSIC)	N/A (AI)	N/A (WI)		
124.039186670145_21.1970275229357	L (BOTH) & A & WIO (INTRINSIC)	Black (LE) Med Imp.	Red (AI) High/Med Imp.	Dark (WI) Med/High Imp.	Black (LI) Low Imp.
261.109154967837_208.020899860918	A & WIO (INTRINSIC)	Red (AI) Med/Low Imp.	Dark (WI) Low/Med Imp.		
233.078104822626_161.866692307692	A, WIO (INTRINSIC), & L (EXTRINSIC)	Black (LE) Med Imp.	Red (AI) Low Imp.	Dark (WI) Low Imp.	
239.127338481013_208.057894101877	A (INTRINSIC) & L (EXTRINSIC)	Black (LE) Med Imp.	Red (AI) Low Imp.		
281.246635949781_342.686808553971	B (INTRINSIC) & L (EXTRINSIC)	Black (LE) Low Imp.	Yellow (BI) Low Imp.		
150.976626042966_17.9383518518518	L (BOTH) & WIO (INTRINSIC)	Black (LE) High Imp.	Black (LI) High Imp.		
485.260099852049_110.452298798798	L (INTRINSIC) & WIO (EXTRINSIC)	Dark (WE) High Imp.	Black (LI) Low Imp.		
274.273285915338_238.389940615058	L (INTRINSIC) & WIO (EXTRINSIC)	N/A (WE) Low Imp.	Black (LI) Low Imp.		
327.133476192115_93.04185042735	L (INTRINSIC) & WIO (EXTRINSIC)	N/A (WE) Low Imp.	Black (LI) Low Imp.		
245.127930082456_129.495900252525	L (INTRINSIC) & WIO (EXTRINSIC)	Dark (WE) High Imp.	N/A (LI)		
316.283770383059_347.79709065934	L (INTRINSIC) & WIO (EXTRINSIC)	Dark (WE) Low/Med Imp.	N/A (LI)		
617.474016589033_429.025435506242	B (BOTH) & WIO (EXTRINSIC)	Blue (BE) Med/High Imp.	Yellow (BI) Low Imp.	N/A (WE) Low Imp.	
211.096155486837_161.846316788321	WIO (INTRINSIC) & B (EXTRINSIC)	Yellow (BE) High Imp.	Dark (WI) Low Imp.		
	*Looking at the Top 15 Attributes: Any w/ NA will be excldued for these discussions				

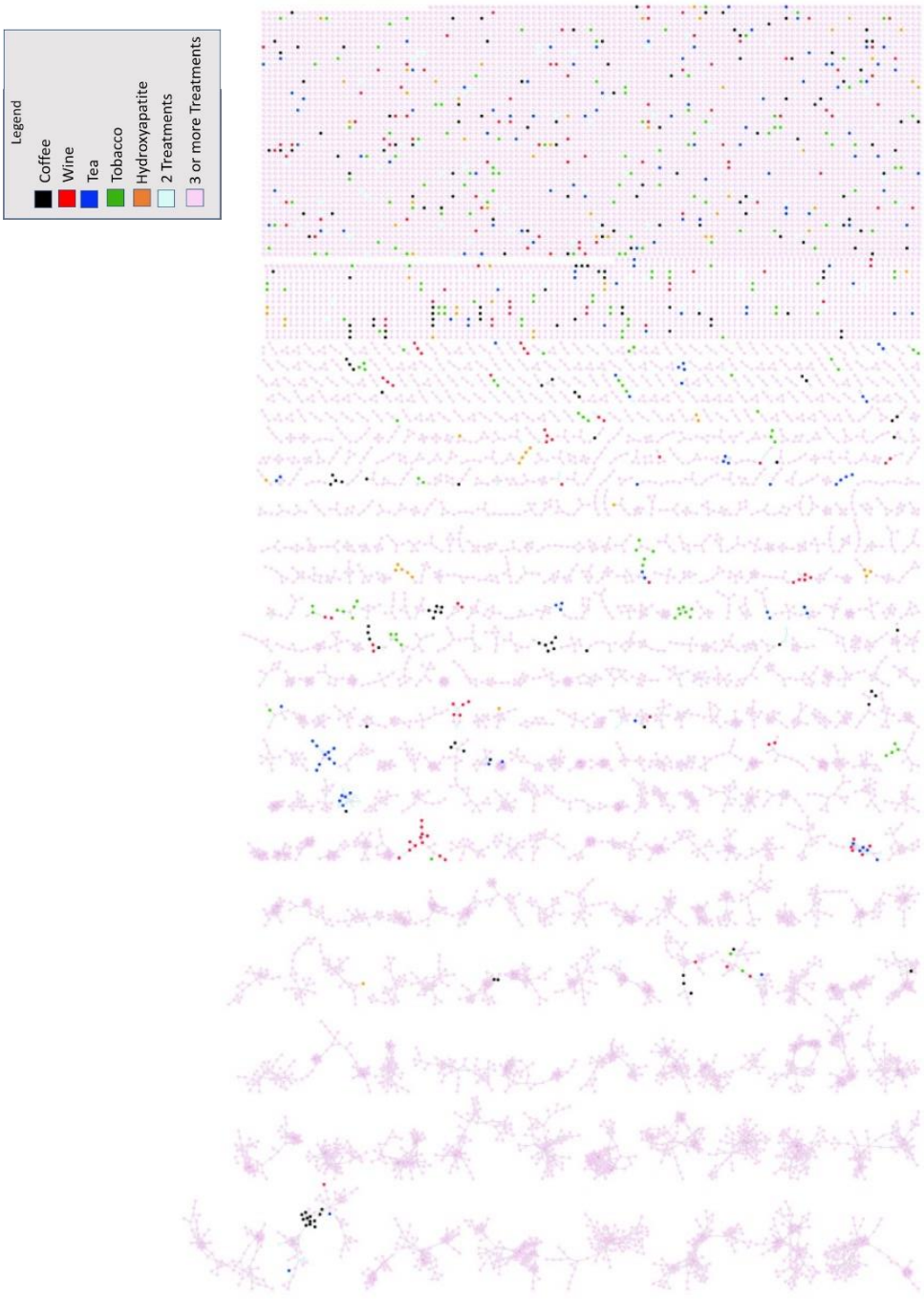
Random Forest allowed for the identification of features that contribute specifically to a particular classification for each color scale, as well as the assessment of accuracy of each classification. The molecular network further permits the identification of each metabolite to be assessed within the GNPS database and compared to molecules of similar composition. It will also allow for the metabolites to be compared based on each staining treatment that took place.

Supervised analysis allowed for discovery of specific compounds that hold key correlations to the L scale and extrinsic staining. Though it is not possible to readily identify these compounds, it is possible to explore the global molecular network to more closely investigate the chemistry of staining.

Molecular networks are visual displays of the chemical space present in tandem mass spectrometry (MS/MS) experiments. It represents similarities between MS/MS spectra as a network of connected nodes. Since structurally similar compounds tend to produce similar MS/MS patterns, this visualization approach can highlight sets of spectra from related molecules as connected nodes in molecular networks. This is possible even when the spectra themselves are not matched to any known compounds. This may help resolve a fundamental problem in metabolomics where limited size of existing reference databases lead to low identification/annotation rates (typically only a few percent) in most untargeted analyses. Also, spectra from the same compound across dataset can get collapsed (clustered) into a single consensus spectrum thus reducing redundancy and making data easier to interpret.

The visualization of molecular networks in GNPS represents each spectrum as a node, and spectrum-to-spectrum alignments as edges (connections) between nodes. Nodes can be supplemented with metadata, including library matches or information that is provided by the user, e.g. as abundance, origin of product, biochemical activity, hydrophobicity etc., which can be reflected in a node's size, shape or color[6]. The GNPS job can be found at: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9180cadafd6740aca203849daed3d90b.

Figure 18. Full molecular network obtained on over 1000 samples.
Single nodes on the right indicate unique consensus MS/MS ions. To be linked together nodes must a similarity of at least 0.66. In the present plot, each colored node represents a treatment type of human teeth


Legend
Coffee
Wine
Tea
Tobacco
Hydroxyapatite
2 Treatments
3 or more Treatments

The legend for *Figure 18* is described below:

Red representing the control group, Yellow representing the stained group, Orange representing the hydroxyapatite group, Green representing the normal group, Purple representing the solvent group, Blue representing the food group, Gray representing 2 or more of the treatment types appearing, and teal is 3 or more treatment types appearing in the node. The nodes represented are not shown through any specific dimensions, but a space in a group. The nodes to the right represent individual nodes not linked to any other clusters due to low similarity to any other nodes.

With an overview of the specific clusters that relate to each staining solution, it is important to better understand the compounds connecting these networks. Network Annotated Prorogation (NAP) allows for investigation of each specific compound. It provides *in silico* predictions that can relate why these compounds have undergone the transformations that take place.

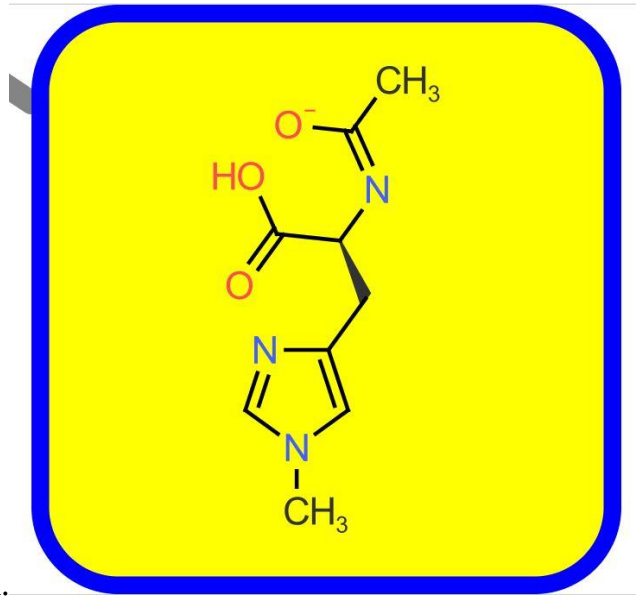
Investigating nodes and clusters of specific compounds can provide important information on how specific sample types are chemically linked. In this study, the key interest is the chemistry of compounds in staining agents. As described above, chemical similarities are represented by connections within the network. Below we explore clusters of compounds that are stain-specific clusters to explore their role in enamel staining in human teeth.

The figures represent Network Annotated Propagation (NAP). The NAP utilizes computationally *in silico* predicted structures and merges this information with

library matches annotations to propose structures for unannotated nodes. After indexing the major staining compounds from tea, coffee, wine, and tobacco they were compared to those found from the most influential metabolites in the study. The link to the NAP generation can be found: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=58d42f481d3d4515b99a4d799c5c30a3.

The metabolites chosen were based on their variance of importance scores from random forest analysis. The metabolites within *Table 1* highlighted are among the top contributors to staining. Metabolites 124.03918667 and 150.976626 were too small for *in silico* generation. The compounds boxed in blue are unannotated within GNPS. The compounds in green are annotated by GNPS. The compound in yellow is the compound discussed for the particular figure.

a.



b.

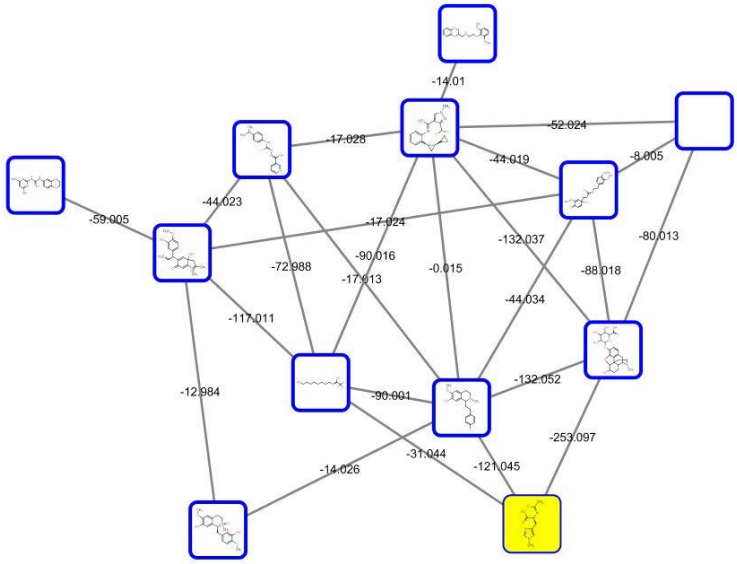
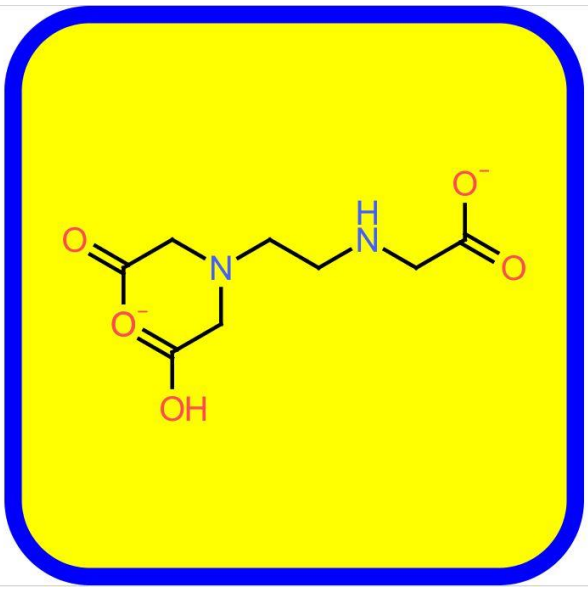


Figure 19. NAP *in silico* prediction of metabolite 211.096155486837 (Cluster ID: 3215) a. The highlighted metabolite prediction containing an aromatic structure and carboxylic group. b. The related compounds to this structure are found within its cluster. Compounds missing in a box are due to a lower level of *in silico* prediction being necessary.

a.



b.

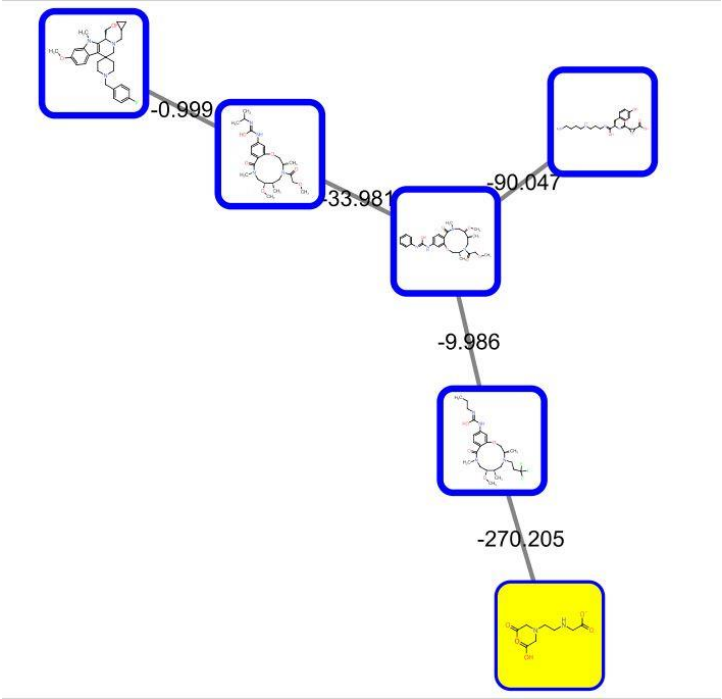
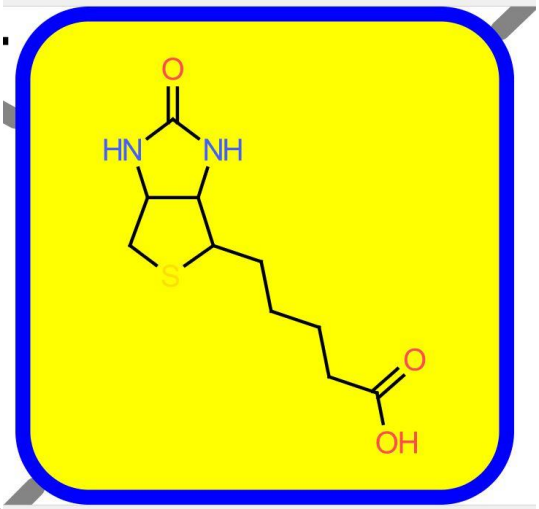


Figure 20. NAP *in silico* prediction of metabolite 233.078104822626 (Cluster ID: 7060) a. The highlighted metabolite prediction containing deprotonated carboxylic group and aromatic structure. b. The related compounds to this structure are found within its cluster.

a.



b.

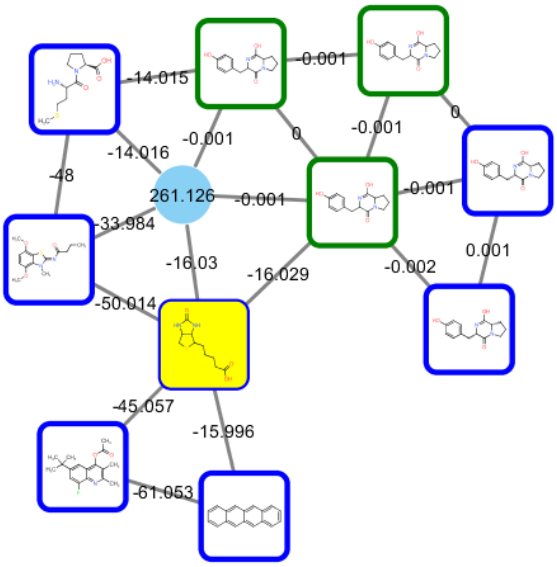
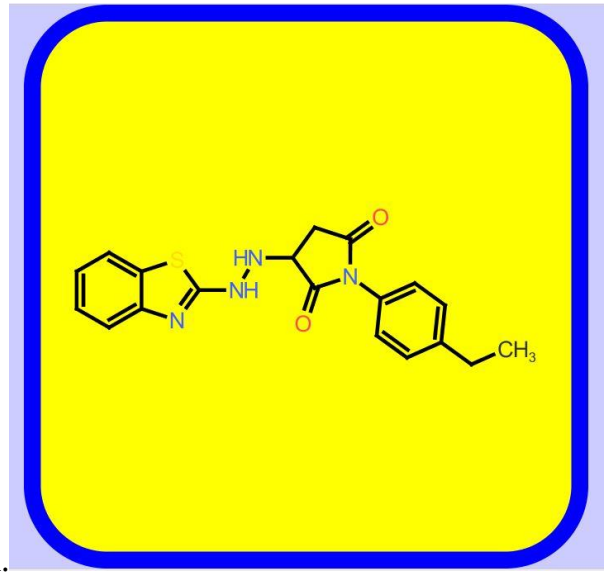


Figure 21. NAP *in silico* prediction of metabolite 245.127930082456 (Cluster ID: 10008) a. The highlighted metabolite prediction containing hemiacetal ketones and a carboxylic branch. b. The related compounds to this structure are found within its cluster. The compounds in green boxes are found within the GNPS database and includes variations of cyclo (tyrosyl-prolyl). The blue node has no known annotations related to its structure.

a.



b.

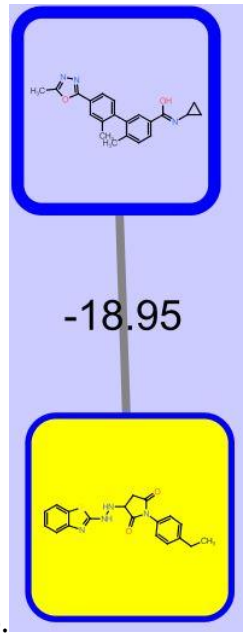
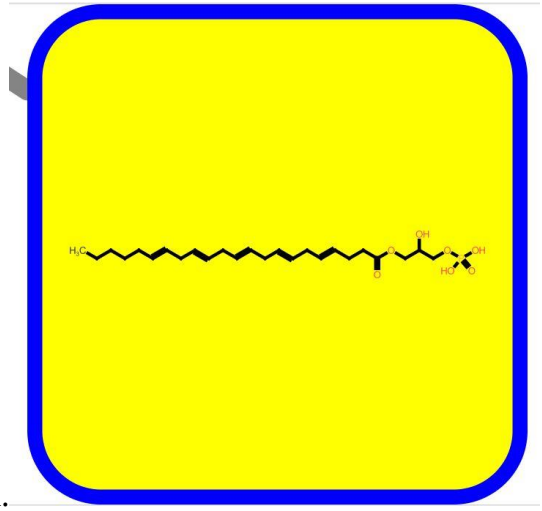


Figure 22. NAP *in silico* prediction of metabolite 367.172219947675 (Cluster ID: 44951) a. The highlighted metabolite prediction containing several aromatic rings and hemiacetal ketones. b. The related compounds to this structure are found within its cluster.

a.



b.

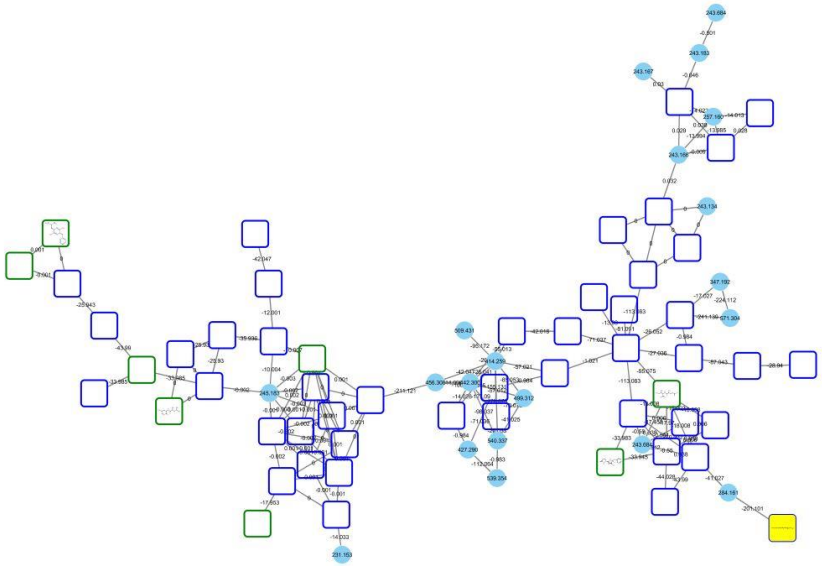


Figure 23. NAP *in silico* prediction of metabolite 485.260099852049 (Cluster ID: 66409) a. The highlighted metabolite prediction containing a phosphor-head group with a 25 carbon fatty chain. b. The related compounds to this structure are found within its cluster. This cluster has a large network of derivations making it complicated to annotate. Closely related within the GNPS database include Glu-Ile and SpectralMatchtoPyroGlu-PhefromNIST14. Compounds missing in a box are due to a lower level of *in silico* prediction being necessary.

After confirming the possible chemical structures for each metabolite that had a high correlation to staining, microbe to metabolite vectorization was the next step in the process. Microbe to metabolite vectorization (MMVEC) allows the inference of interactions across multiple omics datasets. MMVEC estimates the conditional probability that each molecule is present given the presence of a specific microorganism. After ANOVA analysis, *Microbiome Supplementary Figure 4*, batch bias could be found between the three different groups from which the samples came from. Due to this, it was found that only the Rutgers and Therametric samples were significant enough to use for the MMVEC analysis.

The MMVEC analysis provided interesting results about the possible microbes that have a high co-occurrence to the staining correlated metabolites. The figures below provide those visualization for some of the most common microbes found. The downloadable access to this visualization can be found for the Rutgers dataset in supplemental downloads. The downloadable access the Therametric dataset can be found in supplemental downloads. These can be drag and dropped into *view.qiime2.org*.

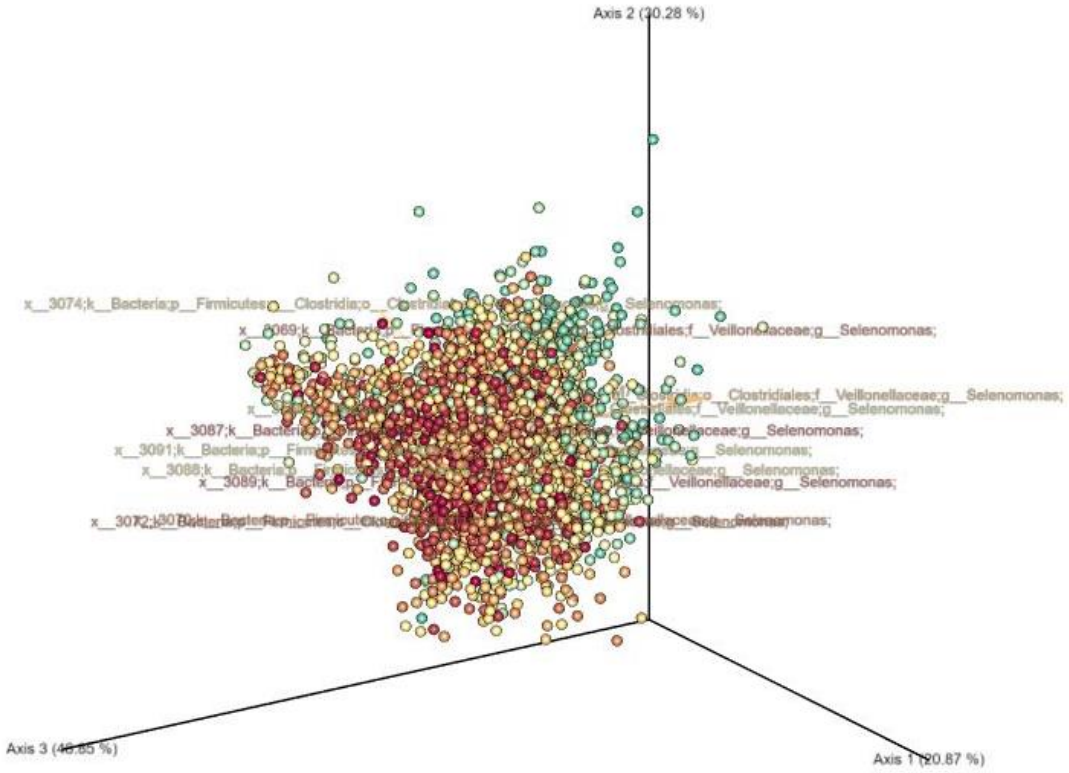


Figure 24. *Selenamonas* within the L scale (Rutgers).
Different *Selenamonas* genus can be found within the ends of the staining scale for black and white. The microbial groups are found pointing towards Axis 1 and away from Axis 3. The L gradient trends away from Axis 2.
**Black is represented by the darker red while the green represents whiter stains.*

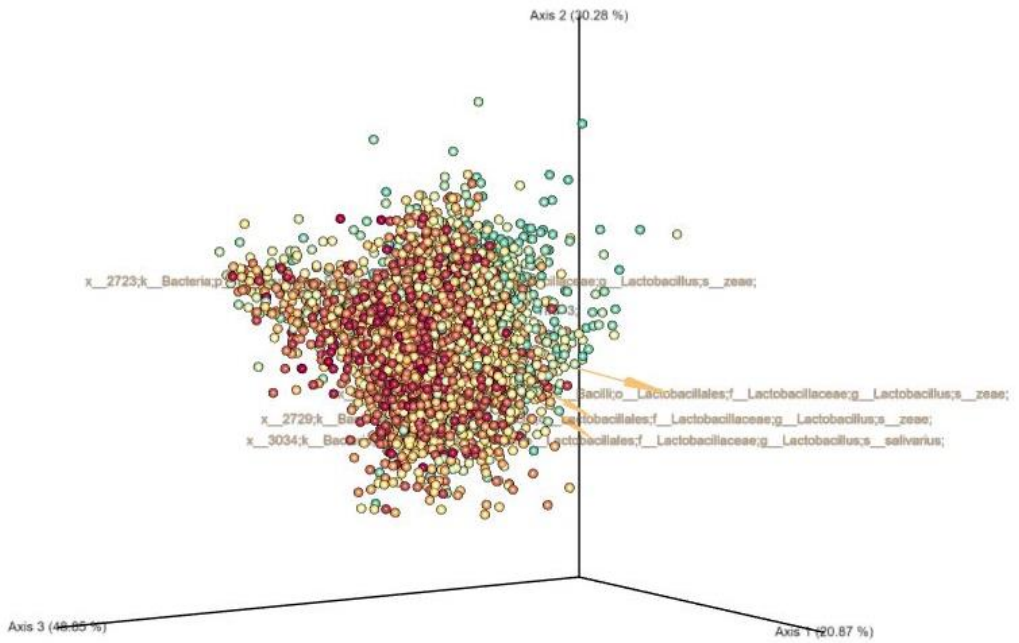


Figure 25. *Lactobacillus zeae and salvarius within the L scale* (Rutgers).
Zeae and Salvarius tend to be found closer to the red points which correlate with blacker stains. The microbial arrows point away from axis 3 in the same direction as the red points. The L gradient trends away from Axis 2.

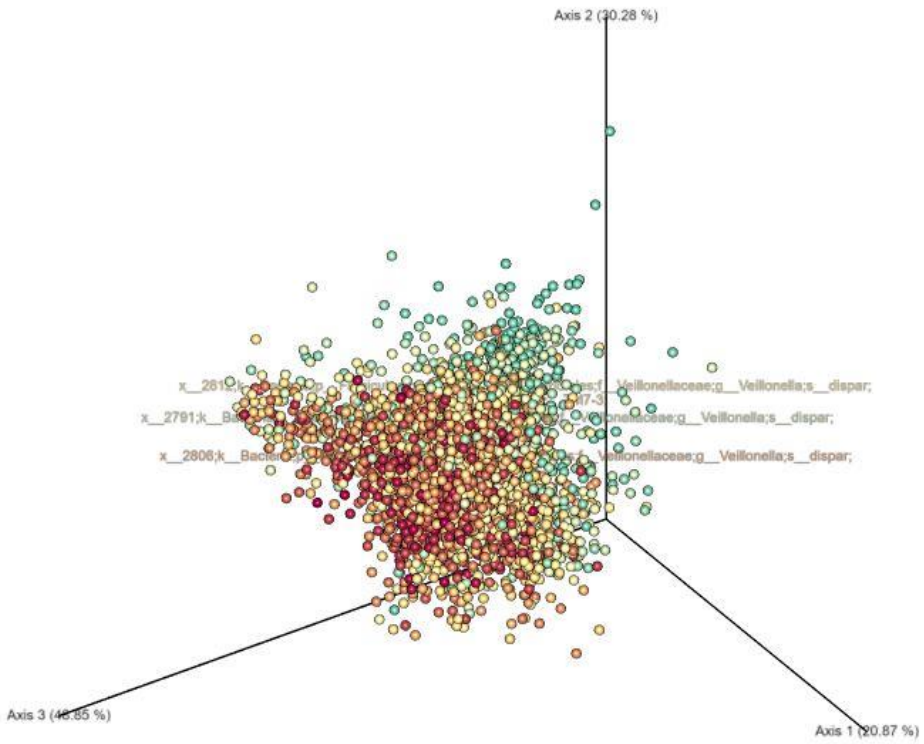


Figure 26. *Veillonella Dispar within the L scale* (Rutgers).
Dispar tend to be found closer to the red points which correlate with blacker stains. Microbial arrows are pointing away from Axis 3. The L gradient trends away from Axis 2.

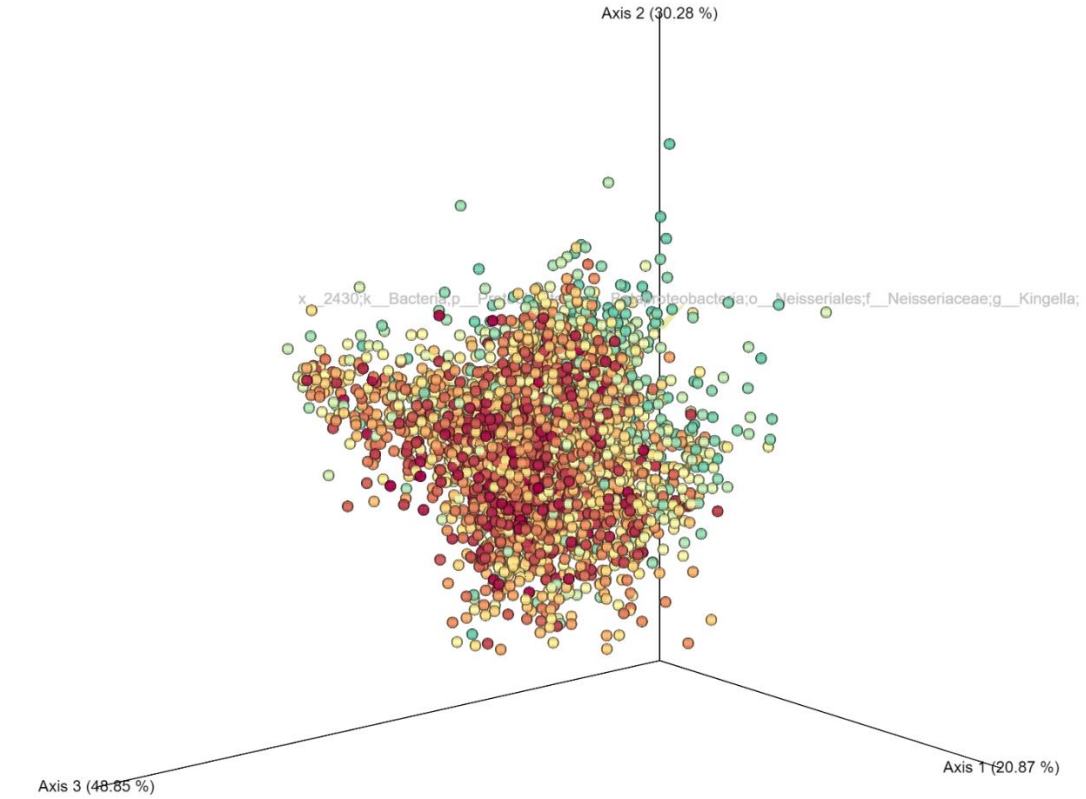


Figure 27. *Kingella within the L scale* (Rutgers).
The *Kingella* genus tends to be found closer to the green points which correlate with whiter stains. The microbial arrow is pointed toward Axis 1. The L gradient trends away from Axis 2.

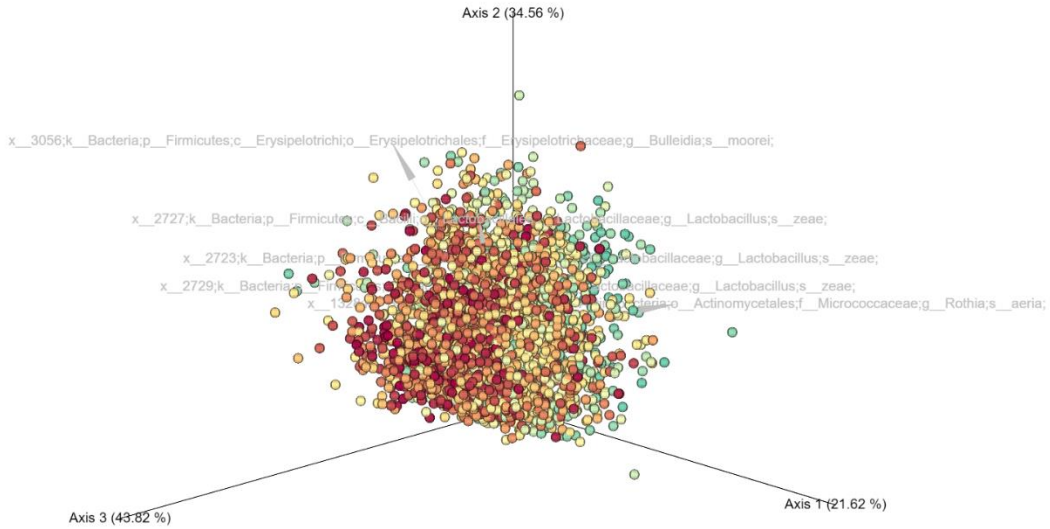


Figure 28. Species within the Therametric dataset on the L scale. *Lactobacillus zeae, Rothia aeria, and Bulleidia moorei (Solobacterium moorei)* all showed to be significant for the metabolites contributing to staining. *Moorei* and *Zeae* have the most influence on black staining while *Aeria* has an impact on white staining.

Within the analysis a co-occurrence ranking table was also designed to correlate the top ten microbes with each metabolite. These tables are found below.

Table 2. Rutgers Top 10 Co-occurrence Microbe List
The microbial community involved with the seven staining microbes all revolve around the same few microbes. These microbes include but are not limited to: *Selenomonas, Lactobacillus zeae, Kingella, Lautropia, Treponema, Leptotrichia*, etc.
**Microbes in orange are not of the oral microbiome or not defined enough to be relevant. Microbes green and bolded are of the oral microbiome.*

featureid	124.039186670145_21.1970275229357
x__3066;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	2.987157664
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	2.953025727
x__2430;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella;	2.524751323
x__2279;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia;	2.452867814
x__953;k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter;	2.335185623
x__2864;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;	2.165624205
x__96;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;	2.000836377
x__3303;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia;	1.880091447
x__3220;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;	1.879649263
x__2440;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;	1.875511168
featureid	150.976626042966_17.9383518518518
x__3074;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	1.041244675
x__2479;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium;s__kroppenstedtii;	1.02131324
x__1699;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;	0.864099401
x__1468;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	0.839941819
x__2689;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;	0.831429798
x__1199;k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Dethiosulfovibrionaceae;g__Pyramidobacter;s__piscolens;	0.788851492
x__2812;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar;	0.786884692
x__2430;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella;	0.770805434
x__1022;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Aliivibrio;	0.746233974
x__1464;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Filifactor;	0.673699764
featureid	211.096155486837_161.846316788321
x__3066;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	**4.613932944**
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	**4.337209789**
x__2430;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella;	**4.00998504**
x__2864;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;	3.379554311
x__2279;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia;	3.374611181
x__2812;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar;	3.226038376
x__953;k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter;	3.181374357
x__96;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;	3.075759635
x__1468;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	3.05134541
x__3303;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia;	2.843618643

featureid	233.078104822626_161.866692307692
x__3066;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	4.170084594
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	4.130729786
x__2430;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella;	3.299391252
x__2864;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;	3.084648658
x__2812;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar;	2.885386826
x__953;k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter;	2.758242622
x__2279;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia;	2.622271044
x__1468;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	2.57085196
x__3034;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__salivarius;	2.566907258
x__96;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;	2.501244654
featureid	245.127930082456_129.495900252525
x__2430;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella;	1.040024535
x__3074;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	0.9339752
x__1699;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;	0.886214615
x__1468;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	0.872033804
x__2440;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;	0.842924578
x__2279;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia;	0.79913155
x__3066;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	0.78692929
x__96;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;	0.782091803
x__3069;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	0.754364905
x__2812;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar;	0.744321971
featureid	367.172219947675_244.442006085193
x__3066;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	3.686982397
x__2430;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella;	3.383745901
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	3.330981657
x__2279;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia;	2.864966167
x__2864;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;	2.689354307
x__2812;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar;	2.639162415
x__96;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;	2.598430318
x__1468;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	2.582133213
x__953;k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter;	2.550432235
x__2440;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;	2.415956578

featureid	485.260099852049_110.452298798798
x__2430;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella;	3.560271539
x__3066;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas;	3.456101063
x__1468;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	2.915053462
x__2812;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;s__dispar;	2.807477811
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	2.746812228
x__2279;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia;	2.74361498
x__96;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;	2.680274211
x__1699;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;	2.659254439
x__2440;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;	2.571439315
x__2864;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;	2.521157464

Table 3. Therametrics Top 10 Co-occurrence Microbe List
The microbial community involved with the seven staining microbes all revolve around the same few microbes. These microbes include but are not limited to: *Bulleidia moorei, pseudomonas, lactobacillus zeae,* etc.
**Microbes in orange are not of the oral microbiome or not defined enough to be relevant. Microbes green and bolded are of the oral microbiome.*

featureid	124.039186670145_21.1970275229357
x__884;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];	1.863640575
x__354;k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;	1.861382824
x__2936;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium;	1.539925913
x__2673;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira;	1.403643019
x__3104;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Schwartzia;	1.273107225
x__3056;k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia;s__moorei;	1.143461908
x__905;k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfovibrionaceae;g__Desulfovibrio;s__mexicanus;	1.039696405
x__2189;k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Geobacillus;	**1.038737434**
x__2230;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	1.022481594
x__911;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;	0.934004054
featureid	150.976626042966_17.9383518518518
x__2936;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium;	2.658791103
x__354;k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;	2.560067962
x__3056;k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia;s__moorei;	**2.549749127**
x__884;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];	2.420598376
x__2673;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira;	2.264133986
x__3104;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Schwartzia;	2.194506824
x__1156;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	1.926689061
x__2847;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;	1.475674303
x__1986;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;	1.461506064
x__1911;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];	1.41887327
featureid	211.096155486837_161.846316788321
x__2391;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia;	2.267639952
x__264;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;	1.738887789
x__1886;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Desulfosporosinus;s__meridiei;	1.732504251
x__2843;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;	1.568506945
x__3056;k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia;s__moorei;	1.518993935
x__1712;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;	1.514835206
x__884;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];	1.504914997
x__1614;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;	1.494367561
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	1.463680239
featureid	233.078104822626_161.866692307692
x__2391;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia;	3.106335816
x__264;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;	2.400820604
x__1886;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Desulfosporosinus;s__meridiei;	2.344487925
x__884;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];	2.293407137
x__3056;k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia;s__moorei;	2.275335855
x__2843;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;	2.240670892
x__2230;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	2.105294019
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	2.08738459
x__1614;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;	2.07513532
x__1712;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;	2.040405841

featureid	245.127930082456_129.495900252525
x__1156;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	0.721204232
x__1618;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;	0.598915167
x__2344;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;	0.529120362
x__1986;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;	0.52014428
x__1886;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Desulfosporosinus;s__meridiei;	0.495833337
x__1201;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	0.479806976
x__1022;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Aliivibrio;	0.471439873
x__2650;k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus;	0.471170372
x__2389;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia;	0.46635179
x__1368;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;	0.445737368
featureid	367.172219947675_244.442006085193
x__2391;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia;	1.710207413
x__3056;k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia;s__moorei;	1.67182913
x__264;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;	1.428498921
x__1886;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Desulfosporosinus;s__meridiei;	1.407559288
x__884;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];	1.400114225
x__2843;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;	1.28761785
x__1156;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	1.273367149
x__1201;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	1.203083993
x__2230;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;	1.154165302
x__2727;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__zeae;	1.138382644
featureid	485.260099852049_110.452298798798
x__1328;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Rothia;s__aeria;	0.98895825
x__1886;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Desulfosporosinus;s__meridiei;	0.974848002
x__2391;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia;	0.938010579
x__1507;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;	0.81290571
x__2882;k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__Gemella;	0.805440174
x__264;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;	0.79798655
x__1201;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	0.754533185
x__1156;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	0.721260855
x__3130;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;	0.702051498
x__3023;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;	**0.698477014**

After exploring the Top Co-Occurance Microbes for both datasets. Further investigation was done to explore the top anti-correlates of those. These microbes only came from the Rutgers Dataset due to the low relevance of the Therametrics Dataset.

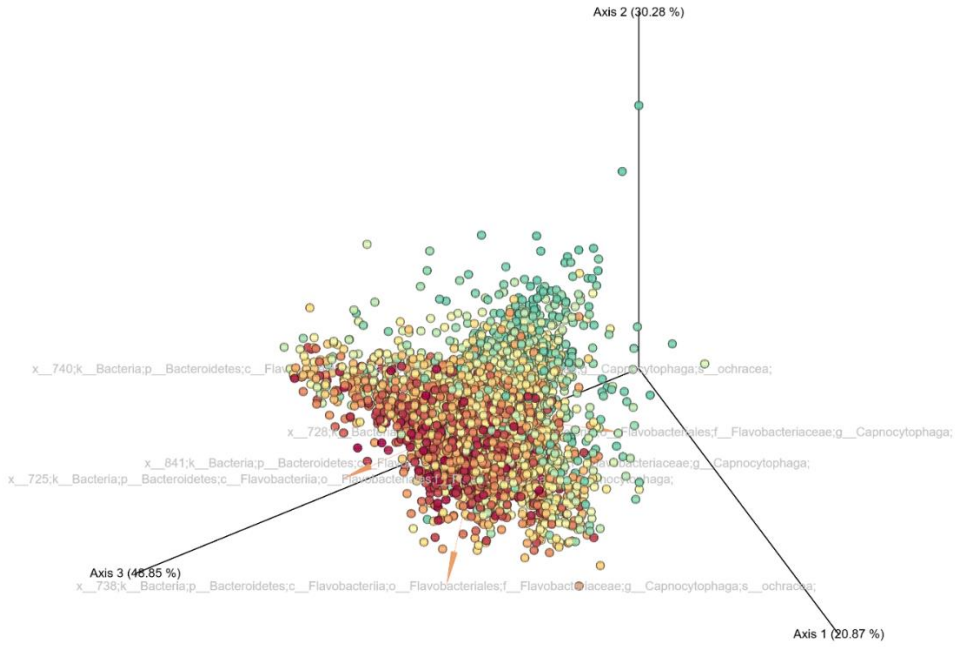


Figure 29. *Capnocytophaga* and *C. Ochracea* within the L scale (Rutgers) *Capnocytophaga* and *C. Ochracea* significantly correlated for the metabolites contributing to staining. Both highly contributed to black staining and show orange arrows pointed away from Axis 3. The L gradient is going towards Axis 2 from black staining to whitening.

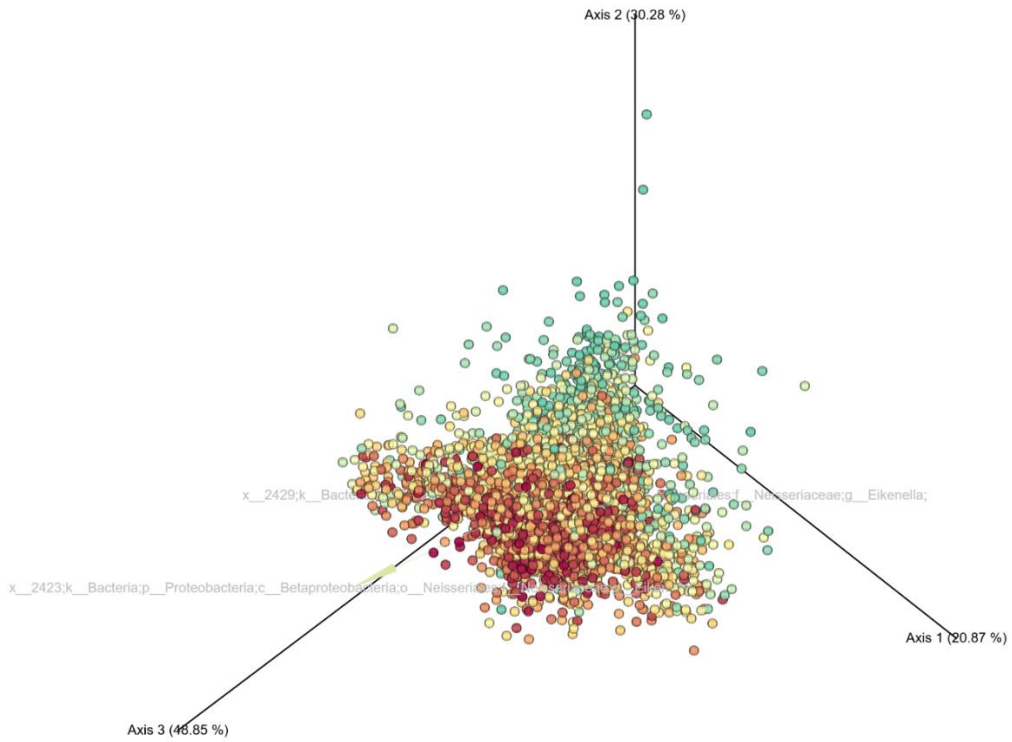


Figure 30. *Eikenella* within the L scale (Rutgers)
The Eikenella spp. significantly correlated for the metabolites contributing to staining. It highly contributes to black staining and show a green arrow pointed away from Axis 3. The L gradient is going towards Axis 2 from black staining to whitening.

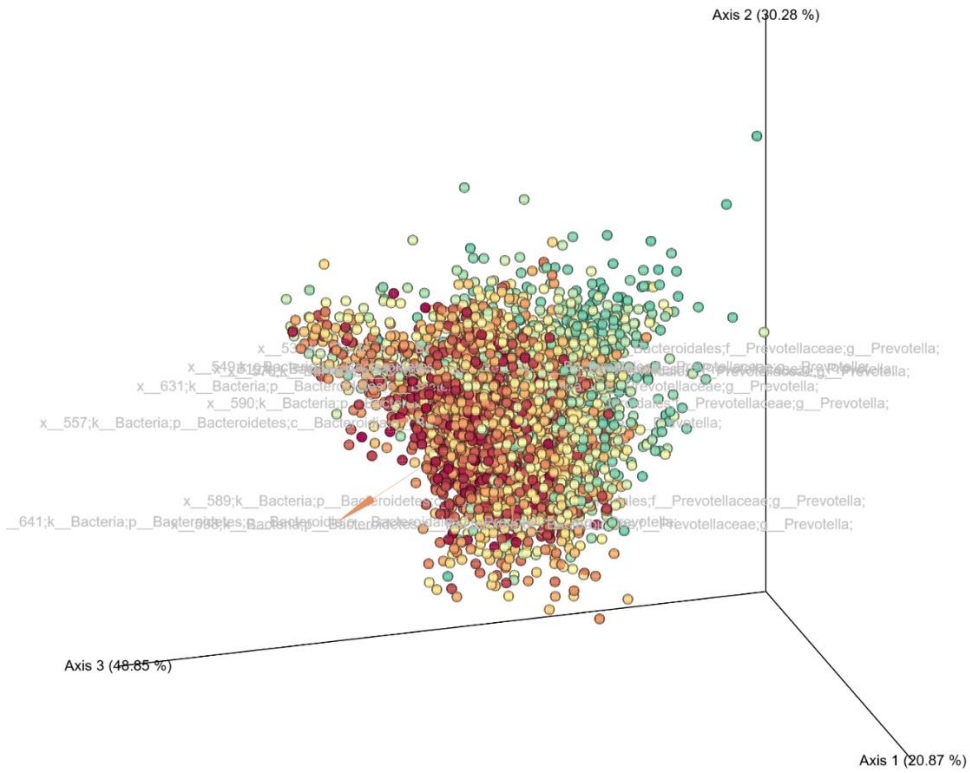


Figure 31. *Prevotella spp.* within the L scale (Rutgers)
The Prevotella spp. significantly correlated for the metabolites contributing to staining. It highly contributes to black staining and shows an orange arrow pointed away from Axis 3. The L gradient is going towards Axis 2 from black staining to whitening.

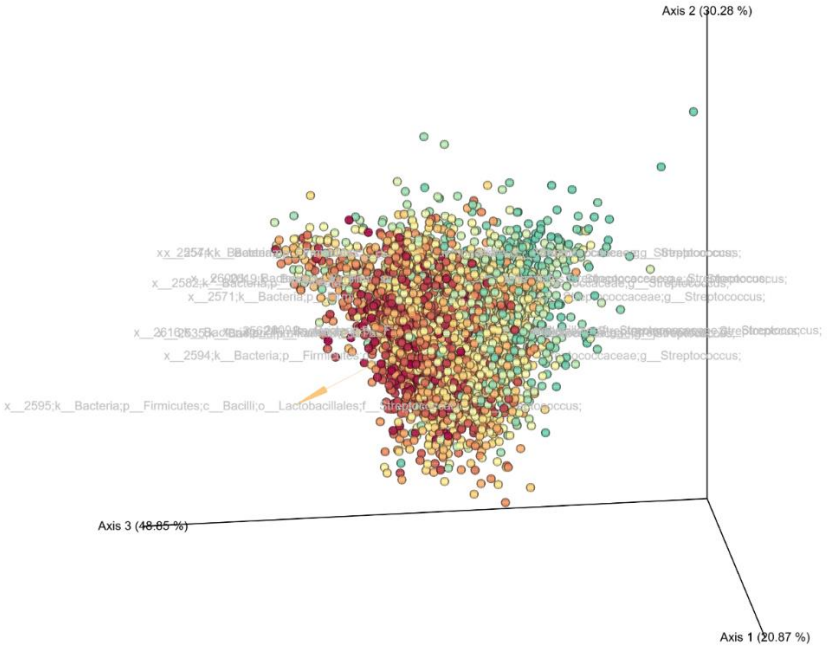
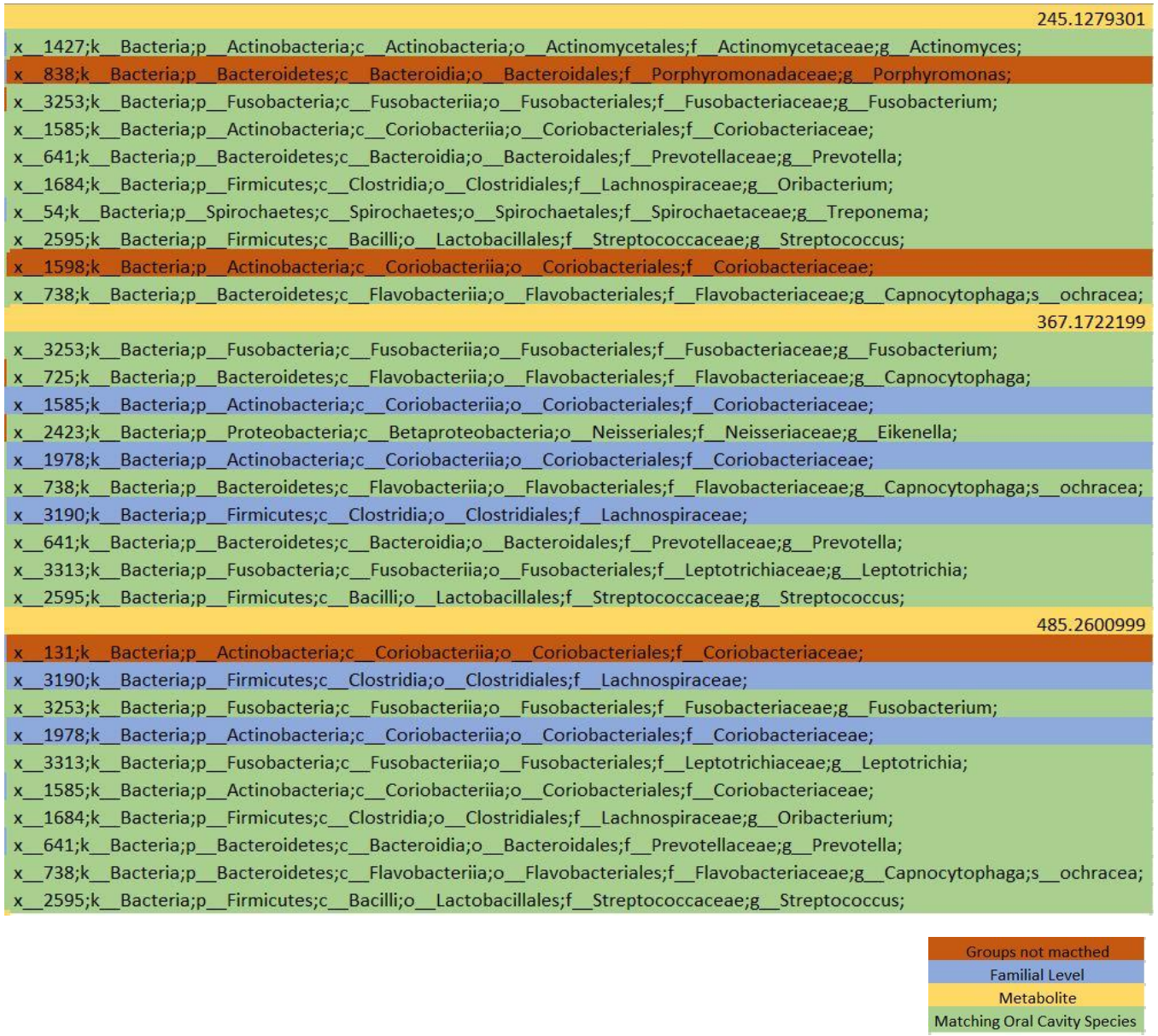


Figure 32. *Streptococcus spp.* within the L scale (Rutgers)
The Streptococcus spp. significantly correlated for the metabolites contributing to staining. It highly contributes to black staining and shows an orange arrow pointed away from Axis 3. The L gradient is going towards Axis 2 from black staining to whitening.

Table 4. Staining Anti-Correlates 10 Top Co-occurrence Microbe List (Rutgers) The microbial community involved with the seven staining microbes all revolve around the same few microbes. These microbes include but are not limited to: *Treponema, Capnocytophaga, Prevotella, Eikenella, Streptococcus, Leptotrichia, Actinomyces, etc.*

124.0391867
x_1978;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x_1654;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x_2996;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus;
x_725;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;
x_107;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;
x_641;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x_3190;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x_2423;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Eikenella;
x_2595;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;
x_3313;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia;
150.976626
x_641;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x_598;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x_1978;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x_1041;k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Cardiobacteriales;f__Cardiobacteriaceae;g__Cardiobacterium;
x_1684;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Oribacterium;
x_54;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;
x_1427;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces;
x_3253;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;
x_2595;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;
x_738;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;s__ochracea;
211.0961555
x_1654;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x_107;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;
x_725;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;
x_738;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;s__ochracea;
x_2423;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Eikenella;
x_1978;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x_641;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x_3190;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x_2595;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;
x_3313;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia;
233.0781048
x_107;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;
x_1003;k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfomicrobiaceae;
x_1654;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x_725;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;
x_2423;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Eikenella;
x_1978;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x_641;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x_2595;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;
x_3190;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x_3313;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia;
Groups not macthed
Familial Level
Metabolite
Matching Oral Cavity Species


245.1279301
x__1427;k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces;
x__838;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas;
x__3253;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;
x__1585;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x__641;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x__1684;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Oribacterium;
x__54;k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae;g__Treponema;
x__2595;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;
x__1598;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x__738;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;s__ochracea;
367.1722199
x__3253;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;
x__725;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;
x__1585;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x__2423;k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Eikenella;
x__1978;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x__738;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;s__ochracea;
x__3190;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x__641;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x__3313;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia;
x__2595;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;
485.2600999
x__131;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x__3190;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;
x__3253;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;
x__1978;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x__3313;k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Leptotrichia;
x__1585;k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;
x__1684;k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Oribacterium;
x__641;k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;
x__738;k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga;s__ochracea;
x__2595;k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;
Groups not macthed
Familial Level
Metabolite
Matching Oral Cavity Species

Chapter IV

Discussion

The unsupervised analysis with Principal Coordinate Analysis (PCoA) allows visualizing multidimensional data by plotting variance within the data in accordance to a specific metric in reduced number of dimensions. In the present case, the totality of detected molecular features for all of the samples in the dataset is being visualized, where a single point on the plot represents one sample and the closeness of two samples in space represents greater similarity of their corresponding molecular distributions. This analysis can be conducted for all detected compounds (MS1-based) or only those that were selected for MS/MS (MS2-based). In the latter case less information is available, as only a subset of ions across the entire experiment have been fragmented to obtain MS2. Figure 2a shows an example of such a plot for all features with MS2 using binary Jaccard distance. There is a clear separation between treatment categories and sample types. Although staining solution samples were clustered tightly together, there was no significant overlap found with the actual stained teeth samples (Figure 2a). The likely reason for clustering is the overall lower amount of ions originated from those samples. This indicates that such staining is not entirely representative of *in vivo* staining conditions. In total, ~270 consensus nodes are were found to be overlapping between normal, stained teeth and staining solutions, *Supplementary Figure 1*. Exploratory analysis was also conducted for all the same detected features (that is, MS1-based), not only those with MS2, as shown on *Figures 2b* and *2c*.

As can be seen, when using the Canberra distance, the general trend in separation is observed by treatment category for PC1 and PC2. Stained teeth clearly form a cluster compared to normal teeth. Hydroxyapatite control samples clustered separately and

spread across PC1. Samples of food-derived stains also clustered separately further along the PC1. A similar trend was observed for the Bray distance metric, which is based on comparative presence/absence between groups, with the exception of stained teeth samples to separate further away from the normal teeth. This suggests that the difference in stained vs. non-stained teeth is driven by acquisition of specific compounds rather than changes in abundances of common compounds in two groups.

The same data can be colored by a measure different than the group. In particular, the measures of whiteness are of interest. Plots with coloring according to the L scale (Scale from White to black measuring +L is more white and -L is more black), the A scale (Scale from red to green, +A is more red and –A is more green), the B scale (Scale from yellow to blue, +B is more yellow and -B is more blue), and the WIO scale (Human perception from Light to Dark teeth, +WIO is more light and -WIO is more dark) are shown on *Figures 3 - 5*. The analysis took place for both intrinsic and extrinsic staining to verify if any significant trends could be seen between the two different surfaces. The trend that is most obvious and apparent is for the L scale. This can be seen through the stratification of the samples from PC3 axis to PC1 axis, which have a gradient corresponding to the L scale values could be observed when using Canberra distance metric (*Figure 3a and 3b*). Little to no differences were observed between the extrinsic and intrinsic plots. This suggests that there's no difference in compounds that are driving staining differences between the enamel and dentin.

The unsupervised analysis strongly indicates chemical differences in the stained teeth compared to normal. Further investigation into which compounds are possibly

responsible for these differences and underlying chemistry captured by the LC-MS/MS using supervised approaches.

The supervised analysis utilizes the information of a priori delineated group to define the separation between them. We have conducted this analysis using the Metaboanalyst platform[10]. Partial Least Squares Discriminant Analysis (PLS-DA) [K] was used to determine which metabolites drive the differences among the experimental categories of interest. Random Forrest (RF) ensemble learning method, a different approach was used to further independently verify the validity of determined discriminating features.

The supervised analysis using Partial least squares Discriminant Analysis (PLS-DA) has been carried out for the categories of "Normal" vs. "Stained" teeth outlined by the metadata (*Figure 7*). These groups were found to be clearly distinct and the separation very robust as evidenced by a very high Q2 value of 0.93103. The Q2 value relates to predictive power of the model to assign samples to the groups and high values that are close to 1 are indicative of substantial differences among groups. The features sorted by the variable importance projection (VIP) scores and thus contribute to the discrimination of these groups are shown on *Figure 7b* for example. An independent classification using the ensemble learning approach, random forest (Figure 7c) has been also conducted. A number of the top features driving the classification do overlap with the features identified by PLS-DA. The Out Of the Bag (OOB) classification error of 0.00921 for random forest analysis is strongly indicative of major differences among these two groups.

An additional category of interest for the staining is the L Scale of whiteness. Correlation analysis using Spearman (*Figure 8, 9*) rank correlations have produced a number of features that largely overlap, but also contain a number of the top features in *Table 1* associated with the "Stained" group from the supervised analyses (*Figure 7a-c*). This serves as an independent validation that the compounds that are negatively associated with whiteness (lower L score) are the same compounds that result in teeth to be classified as "Stained". The chemical identity of these compounds was explored using molecular networking and will be discussed in later materials.

The supervised analysis revealed that there are several features that clearly discriminate the stained teeth. Interestingly, most of these features were also found to have strong correlation with the L scale measurements which further implicates these compounds to staining. Unfortunately, most of the features in *Table 1*, including the most discriminant feature from PLS-DA analysis, m/z 599.247221578329 are currently unidentified. To explore the chemical identities of the staining chemicals, molecular networks were explored.

Due to the success of the PLS-DA analysis, exploring Spearman correlations for all of the color scales were analyzed. Few matches could be seen between extrinsic and intrinsic staining with the different color scales that were explored. Looking at *Table 1*, metabolite matches were seen in the L extrinsic color scale. This table was also compared to the random forest analysis. Due to the relatively small amounts of metabolite correlations that have occurred, it is inconclusive to state that there is a major effect of the metabolites on intrinsic staining. The random forest data provided validity

of the characteristics that were defined of each tooth within the study. Looking at *Table 1*, most metabolite matches were seen in the L extrinsic color scale and the WIO perception scale. This hypothesizes that these metabolites may have a correlation with the black staining (negative L scale) on teeth and their perception on the human eye (WIO) to a darker tooth, rather than any effects from the A and B scales. This discovery leads back into the key metabolites that had significant impact on the study.

Metabolites 124.0391866701, 150.9766260429, and 233.078104822626 were found to have the most significance. Metabolites 124.0391866701 and 150.9766260429 were too small for *in silico* prediction so there is no predicted structure to analyze. Metabolite 233.078104822626 had a predicted structure similar to biotin (Vitamin B7). Since this is a *in silico* prediction there is no conclusive data to support this. Therefore, MS1 data would need to be analyzed to observe possible fragmentation patterns. These compounds were analyzed and assessed by NAP. Studying the global network there were many compounds that were defined from different treatments types used. However, none of the compounds were directly related to one specific staining solution.

Due to newly developed software, microbe to metabolite vectorization (MMVEC), enables significant contributions to this study. This analysis directly correlates specific microbes and specific metabolites to metadata features that may be of interest. MMVEC has shown a significantly more accurate analysis with the correlations of metabolites than any other correlation methods previously developed[15]. Due to batch bias there we two sets of data that were explored. The correlation to each

color scale provided the basis of directionality. Color scales A and B can be found in *Supplementary Figures 2-11.*

There were four key genus that were involved along with three key species. The four genus include *Laurtropia, Kingella,* and *Selenomonas*. The three key species level microbes include *Lactobacillus zeae*, *Veillonella dispar*, and *Bulledia moorei* (*Solobacterium moorei*). These microbes were of particular interest because they all had a high co-occurrence with each other the metabolites secreted that contributed highly to whitening and their relation to periodontal disease. The metabolites may have been produced from chemical reactions or secreted from a particular microbe or group.

The gram negative, facultative anaerobe *Lautropia* has only one known oral species, which is *L. mirabilis*[16,17]. *L. Mirabilis* is known to be involved with subgingival gingivitis, basic periodontal health, and subgingival biofilms[18,19]. This interaction could serve as one of the earliest colonizer when stain vulnerability is develops.

The gram-negative, anaerobic, *Kingella* genus (*Figure 27*) includes 6 species that are found within the oral cavity[17]. *Kingella oralis* has been found to be involved with healthy and diseased patients of periodontal disease[20]. This implies th*at Kingella* could be another founding member of this colonization.

The gram negative, anaerobic, *Selenomonas genus* (*Figure 24*) includes 25 species that are found within the oral cavity. *Selenomonas noxia* is a known periodontal

disease bacteria[21]. *S. sputigena* has a generalized role in aggressive periodontal disease [22].

The gram-positive, anerobic *Lactobacillus zeae* later classified as *L. casei* and the facultative anaerobe *L. salivarius* (*Figure 25*) are both found within the oral cavity along with 24 other species[25, 26]. *L. casei* is found to be mechanically stimulated by saliva and found within biofilm formation[27]. The microbe was also found in patients with Amelogenesis imperfecta protecting against dental caries. This could imply that a staining environment is independent of the caries process in general, or under specific conditions like the Amelogenesis imperfecta environment[28]. *L. Salivarius* has been involved in controlling periodontitis, dental caries and malodor. This could perhaps be the microbiome fighting back and recruiting specific microbes to fight off the disease. It was also involved in effecting epigallocatechin gallate which is a confirmed active compound in the tea stained samples. This could be the microbe breaking down or re-constructing the compound into a staining agent [29].

The gran negative, anaerobic, *Viellonella dispar* (*Figure 26*) is one out of 7 within the species. *V. Dispar* has been known to have early colonizing interactions with *Streptococcus* spp. through quorum sensing[31]. Often being involved in soft tissues infections it would make sense to be a microbe involved in the spectrum of periodontal disease[31]. It also implies that *V.Dispar* may contribute to a healthy community fighting back against periodontal disease and other types of infections. They commonly inhabit the tongue, carious lesions, and saliva of caries-free individuals [32].

The gram-positive, obligate anaerobic, *Solobacterium moorei*, previously known as *B. moorei*, is the only oral species within *Solobacterium* genus (*Figure 28*). The species is highly correlated to Halitosis, bad breath [33]. The species has been known to be reduced by chlorohexidine washes, which may contribute to less staining if used prior to large amounts of plaque build-up, tongue-coating, and gingivitis[34]. The coaggregation of *S.moorei* appears to be a possible commensal bacteria to disease within the Therametrics dataset. However, in comparison to the Rutgers dataset, *S.moorei* was ranked below the top fifty microbes co-occurring. This makes the microbe inconclusive to be a member of the whitening microbes.

The Rutgers study appears to give the best depiction of what the staining microbial community could be comprised of. Investigation into the species level for one microbe and genus level for nine of the microbes within the staining community. This includes: *Capnocytophaga Ochracea, Eikenella, Prevotella, Streptococcus, Actinomyces, Fusobacterium, Leptotrichia, Oribacterium, and Treponema,*

The gram-negative, facultative anaerobe *Capnocytophaga Ochracea,* is one of twenty-two *Capnocytophaga spp.* known in the oral cavity[17]. *C. Ochracea* is a known periodontal pathogen and has been known to be involved with most of the genus involved in the periodontal disease microbial community. This microbe has also been known to be involved in halitosis [36]. This could be the healthy microbes fighting off these pathogens and release foul smells from their death.

The gram-negative, anaerobes *Eikenella spp.* has two known oral species[17]. The species has been found in localized primary dentition in a case report[35]. Further

investigation into this species would provide significant insight into this microbial community study. This could imply that the species provides a key function in the dysbiosis of a non-staining environment.

The gram-negative, obligate anaerobes *Prevotella spp.* has fifty-one known oral species[17]. *Prevotella* is known to be involved with halitosis and periodontal disease communities[36]. This could imply that *Prevotella* is a main reason for staining and should be investigated for species level interactions to test metabolism.

The gram-positive, facultative anaerobes *Streptococcus spp.* has thirty-nine known oral species[17]. *S. mutans* is a species known to cause dental caries, it would be interesting to know whether or not this specific species was involved with tooth staining as well. It would be interesting to know if either dental caries and periodontal disease happen independent of each other and recruit these specific microbes to mediate unhealthy and healthy environments.

The gram-positive, facultative anaerobes *Actinomyces spp.* has thirty known oral species[17]. Multiple species within this genus are known to be normal members of a healthy oral environment. It would be interesting to specifically narrow down a species to see if this was a normal commensal bacteria or pathogenic and furthering periodontal disease environment in staining conditions.

The gram-negative, anaerobes *Fusobacterium spp.* has fourteen known oral species[17]. The species has been known to appear in late colonization of oral biofilms[39]. Implying that *Fusobacterium* may serve as a spectator when relating to staining.

Exploring a specific species of this bacteria could provide more insight into this implication.

The gram-negative, anaerobes *Leptotrichia spp.* has twenty-one known oral species[17]. The species is known to be found in both healthy environments as well as halitosis[36]. It has also been known to elevate in levels in diabetic patients, and diabetes is known to expedite periodontal disease by complicating the metabolic control[40]. Narrowing down the species would further implicate the relationship of periodontal disease and tooth staining.

The gram-positive, strict anaerobes *Oribacterium spp.* has five known oral species[17]. Little studies have been done with this species, however it has been known to be involved with chronic obstructive pulmonary disease (COPD). The study aimed to connect the group with periodontal disease, however the species decreased in amount, implying it arises to attempt to develop a healthy environment[41].

The gram-negative, aerobic or microaerophilic *Treponema* genus has over 50 different species that can be traced back to the oral cavity[17]. *T. denticola* is a known periodontal pathogen[23]. It is also known to be one of the three microbiota found to cause bone destruction and periodontal disease. This increased effect on the enamel could be one of the main factors involved in a tooth staining environment[24]. *Treponema* has species found both in the whitening and staining communities developing this implication more as well as a staining environment being linked periodontal disease environments.

These microbes lead to particular interest of each metabolite and their chemical make-up through their metabolism. Using cytoscape software, and NAP it was possible to see the potential candidates in relation to the compounds. Specifically these are special due to their association with coffee, wine, tea, or tobacco. These metabolites correlate with previously identified metabolites. However, there may be further investigations that can explain how the metabolites affect the chemistry of the staining solutions. Though *in silico* predictions, these compounds don't appear to be naturally occurring in the oral cavity. Thus, further corroborating the hypothesis that this staining is caused by microbial communities.

Chapter V

Conclusion and Perspective

The future directions look extremely bright with this study due to the development of MMVEC analysis. It provides hope that future studies gain a better understanding of the foods in society that are consumed, and how they are metabolized within microbial communities in the oral cavity. Staining effects appear to be most prevalent in the enamel. The dentin holds similar compounds, however not in high amounts that sway staining. The staining solutions used in the study, do not hold a significant impact in staining to be seen through normal teeth. Periodontal microbial communities appear to have the highest impact in staining. Narrowing these genus down to species level will help to drive this point and further validate this study. The study hopes to determine if similar compounds can be found in other countries and continents as well. It would be interesting to see if these effects were to be influenced due to dental carries, tooth porosity, and erosion – which all have specific effects on compounds that have been known to stain teeth. It would also be interesting to see if any linkage to periodontal disease or gingivitis could be made. Malodor prevalence in relation to these diseases would make for stronger connections between these different groups. It would be advantageous to use a culture microbe to fight off pathogens and potentially revive unhealthy environments, which lead to plague formation, periodontal disease, and tooth staining.

This thesis is coauthored with Aksenov, Alexander, Minich, Jeremiah, and Lejzerowitz, Franck. The thesis author was the primary author of this material.

REFERENCES

1. Susan O. Griffin, Judith A. Jones, Diane Brunson, Paul M. Griffin, William D. Bailey, "Burden of Oral Disease Among Older Adults and Implications for Public Health Priorities", *American Journal of Public Health* 102, no. 3 (March 1, 2012): pp. 411-418.

2. Preshaw, P M; Alba, A; Herrera D.; Jepsen, S.; Konstantinidis; Makrilakis, K.; Taylor, R. l. "Periodontitis and diabetes: a two-way relationship." Diabetologia vol. 55,1 (2012): 21-31. doi:10.1007/s00125-011-2342-y

3. Benjamin, Regina M. "Oral health: the silent epidemic." Public health reports (Washington, D.C. : 1974) vol. 125,2 (2010): 158-9. doi:10.1177/003335491012500202

4. Watts, A, and M Addy. "Tooth Discolouration and Staining: a Review of the Literature." *British Dental Journal*, vol. 190, no. 6, 2001, pp. 309–316., doi:10.1038/sj.bdj.4800959.

5. Loesche WJ. Microbiology of Dental Decay and Periodontal Disease. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 99.

6. Wang, M., Carver, J., Phelan, V. . Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34,** 828–837 (2016). https://doi.org/10.1038/nbt.3597

7. Louis Felix Nothias, Daniel Petras, Robin Schmid, Kai Dührkop, Johannes Rainer, Abinesh Sarvepalli, Ivan Protsyuk, Madeleine Ernst, Hiroshi Tsugawa, Markus Fleischauer, Fabian Aicheler, Alexander Aksenov, Oliver Alka, Pierre-Marie Allard, Aiko Barsch, Xavier Cachet, Mauricio Caraballo, Ricardo R. Da Silva, Tam Dang, Neha Garg, Julia M. Gauglitz, Alexey Gurevich, Giorgis Isaac, Alan K. Jarmusch, Zdeněk Kameník, Kyo Bin Kang, Nikolas Kessler, Irina Koester, Ansgar Korf, Audrey Le Gouellec, Marcus Ludwig, Martin H. Christian, Laura-Isobel McCall, Jonathan McSayles, Sven W. Meyer, Hosein Mohimani, Mustafa Morsy, Oriane Moyne, Steffen Neumann, Heiko Neuweger, Ngoc Hung Nguyen, Melissa Nothias-Esposito, Julien Paolini, Vanessa V. Phelan, Tomáš Pluskal, Robert A. Quinn, Simon Rogers, Bindesh Shrestha, Anupriya Tripathi, Justin J.J. van der Hooft, Fernando Vargas, Kelly C. Weldon, Michael Witting, Heejung Yang, Zheng Zhang,

Florian Zubeil, Oliver Kohlbacher, Sebastian Böcker, Theodore Alexandrov, Nuno Bandeira, Mingxun Wang, Pieter C. Dorrestein bioRxiv 812404; doi: https://doi.org/10.1101/812404

8. da Silva RR, Wang M, Nothias L-F, van der Hooft JJJ, Caraballo-Rodríguez AM, Fox E, . (2018) Propagating annotations of molecular networks using *in silico* fragmentation. PLoS Comput Biol 14(4): e1006089. https://doi.org/10.1371/journal.pcbi.1006089

9. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics Frank Dieterle,Alfred Ross,Götz Schlotterbeck, and, and Hans Senn* Analytical Chemistry 2006 78 (13), 4281-4290 DOI: 10.1021/ac051632c

10. Jasmine Chong, Othman Soufan, Carin Li, Iurie Caraus, Shuzhao Li, Guillaume Bourque, David S Wishart, Jianguo Xia, MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W486–W494, https://doi.org/10.1093/nar/gky310

11. Yoshiki Vázquez-Baeza, Meg Pirrung, Antonio Gonzalez, Rob Knight, EMPeror: a tool for visualizing high-throughput microbial community data, *GigaScience*, Volume 2, Issue 1, December 2013, 2047–217X–2–16, https://doi.org/10.1186/2047-217X-2-16

12. Pluskal, Tomáš, and Matej Orešič. "MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data." *National Institue of Health*, 23 July 2010, doi:10.1186/1471-2105-11-395

13. Zhang, X., and B. A. Wandell. "A Spatial Extension of CIELAB for Digital Color-Image Reproduction." *Journal of the Society for Information Display*, vol. 5, no. 1, 1997, p. 61, doi:10.1889/1.1985127.

14. Wang, Mingxun, . "Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking." *Nature Biotechnology*, vol. 34, no. 8, Aug. 2016, pp. 828–37.

15. Morton, J.T., Aksenov, A.A., Nothias, L.F. . Learning representations of microbe–metabolite interactions. *Nat Methods* **16,** 1306–1314 (2019). https://doi.org/10.1038/s41592-019-0616-3

16. Bik, E., Long, C., Armitage, G. . Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J* **4,** 962–974 (2010). https://doi.org/10.1038/ismej.2010.30

17. Isabel Fernández Escapa, Tsute Chen, Yanmei Huang, Prasad Gajare, Floyd E Dewhirst, and Katherine P Lemon. **New insights into human nostril microbiome from the expanded Human Oral Microbiome Database (eHOMD): a resource for species-level identification of microbiome data from the aerodigestive tract.** DOI: 10.1128/mSystems.00187-18

18. Lim, Y.K., Park, S., Lee, W. . *Lautropia dentalis* sp. nov., Isolated from Human Dental Plaque of a Gingivitis Lesion. *Curr Microbiol* **76,** 1369–1373 (2019). https://doi.org/10.1007/s00284-019-01761-1

19. Soares G.M.S., Faveri M. (2019) Defining the Healthy Oral Microbiome. In: Azcarate-Peril M., Arnold R., Bruno-Bárcena J. (eds) How Fermented Foods Feed a Healthy Gut Microbiota. Springer, Cham

20. Chen, C. "Distribution of a Newly Described Species, Kingella Oralis, in the Human Oral Cavity." *Oral Microbiology and Immunology*, vol. 11, no. 6, 1996, pp. 425–427., doi:10.1111/j.1399-302x.1996.tb00206.x.

21. Cruz, P., Mehretu, A.M., Buttner, M.P. . Development of a polymerase chain reaction assay for the rapid detection of the oral pathogenic bacterium, *Selenomonas noxia*. *BMC Oral Health* **15,** 95 (2015). https://doi.org/10.1186/s12903-015-0071-1

22. Gonçalves, L. F. H., . "Levels OfSelenomonasspecies in Generalized Aggressive Periodontitis." *Journal of Periodontal Research*, vol. 47, no. 6, 2012, pp. 711–718., doi:10.1111/j.1600-0765.2012.01485.x.

23. Verma, Raj K., . "Porphyromonas GingivalisandTreponema DenticolaMixed Microbial Infection in a Rat Model of Periodontal Disease." *Interdisciplinary*

Perspectives on Infectious Diseases, vol. 2010, May 2010, pp. 1–10., doi:10.1155/2010/605125.

24. Bodet C, Chandad F, Grenier D. [Pathogenic potential of Porphyromonas gingivalis, Treponema denticola and Tannerella forsythia, the red bacterial complex associated with periodontitis] Pathologie-biologie. 2007 Apr-May;55(3-4):154-162. DOI: 10.1016/j.patbio.2006.07.045.

25. Dicks, L. M. T., . "Reclassification of Lactobacillus Casei Subsp. Casei ATCC 393 and Lactobacillus Rhamnosus ATCC 15820 as Lactobacillus Zeae Nom. Rev., Designation of ATCC 334 as the Neotype of L. Casei Subsp. Casei, and Rejection of the Name Lactobacillus Paracasei." *International Journal of Systematic Bacteriology*, vol. 46, no. 1, 1996, pp. 337–340., doi:10.1099/00207713-46-1-337.

26. Yang, R., . "Determining the Genetic Diversity of Lactobacilli from the Oral Cavity." *Journal of Microbiological Methods*, vol. 82, no. 2, 2010, pp. 163–169., doi:10.1016/j.mimet.2010.05.010.

27. Inui, Taichi, . "Effect of Mechanically Stimulated Saliva on Initial Human Dental Biofilm Formation." *Scientific Reports*, vol. 9, no. 1, 2019, doi:10.1038/s41598-019-48211-3.

28. Kammoun, Rym, . "Dental Caries and Hypoplastic Amelogenesis Imperfecta: Clinical, Structural, Biochemical and Molecular Approaches." *Microbial Pathogenesis*, vol. 135, 2019, p. 103615., doi:10.1016/j.micpath.2019.103615.

29. Higuchi, Takuya, . "Effects of Lactobacillus Salivarius WB21 Combined with Green Tea Catechins on Dental Caries, Periodontitis, and Oral Malodor." *Archives of Oral Biology*, vol. 98, 2019, pp. 243–247., doi:10.1016/j.archoralbio.2018.11.027.

30. Mashima, Izumi, and Futoshi Nakazawa. "Interaction between Streptococcus Spp. and Veillonella Tobetsuensis in the Early Stages of Oral Biofilm Formation." *Journal of Bacteriology*, vol. 197, no. 13, 2015, pp. 2104–2111., doi:10.1128/jb.02512-14.

31. Actor, Jeffrey K. “Clinical Bacteriology.” *Elsevier's Integrated Review Immunology and Microbiology*, 2012, pp. 105–120., doi:10.1016/b978-0-323-07447-6.00012-0.

32. Do, Thuy, . “Transcriptomic Analysis of Three Veillonella Spp. Present in Carious Dentine and in the Saliva of Caries-Free Individuals.” *Frontiers in Cellular and Infection Microbiology*, vol. 5, 2015, doi:10.3389/fcimb.2015.00025.

33. Vancauwenberghe, Frederique, . “The Role OfSolobacterium Mooreiin Oral Malodour.” *Journal of Breath Research*, vol. 7, no. 4, 2013, p. 046006., doi:10.1088/1752-7155/7/4/046006.

34. Pham, Thuy Anh Vu, and Ngoc Thi Xuan Nguyen. “Efficacy of Chlorine Dioxide Mouthwash in Reducing Oral Malodor: A 2-Week Randomized, Double-Blind, Crossover Study.” *Clinical and Experimental Dental Research*, vol. 4, no. 5, Oct. 2018, pp. 206–215., doi:10.1002/cre2.131.

35. Suzuki, Junji, . “Localized Aggressive Periodontitis in Primary Dentition: A Case Report.” *Journal of Periodontology*, vol. 74, no. 7, 2003, pp. 1060–1066., doi:10.1902/jop.2003.74.7.1060.

36. Ren, Wen, . “Supragingival Plaque Microbial Community Analysis of Children with Halitosis.” *Journal of Microbiology and Biotechnology*, vol. 26, no. 12, 2016, pp. 2141–2147., doi:10.4014/jmb.1605.05012.

37. Acharya, Aneesha . “Species-Level Salivary Microbial Indicators of Well-Resolved Periodontitis: A Preliminary Investigation.” *Frontiers in cellular and infection microbiology* vol. 9 347. 11 Oct. 2019, doi:10.3389/fcimb.2019.00347

38. Durand, R., Roufegarinejad, A., Chandad, F. . Dental caries are positively associated with periodontal disease severity. *Clin Oral Invest* **23,** 3811–3819 (2019). https://doi.org/10.1007/s00784-019-02810-6

39. Leanne M Cleaver, Rebecca Moazzez & Guy H Carpenter (2019) Mixed aerobic-anaerobic incubation conditions induce proteolytic activity from in vitro

salivary biofilms, Journal of Oral Microbiology, 11:1, DOI: 10.1080/20002297.2019.1643206

40. Grossi, Sara G., and Robert J. Genco. “Periodontal Disease and Diabetes Mellitus: A Two-Way Relationship.” *Annals of Periodontology*, vol. 3, no. 1, 1998, pp. 51–61., doi:10.1902/annals.1998.3.1.51.

41. Wu, Xingwen, . “16S RDNA Analysis of Periodontal Plaque in Chronic Obstructive Pulmonary Disease and Periodontitis Patients.” *Journal of Oral Microbiology*, vol. 9, no. 1, 2017, p. 1324725., doi:10.1080/20002297.2017.1324725.

APPENDIX

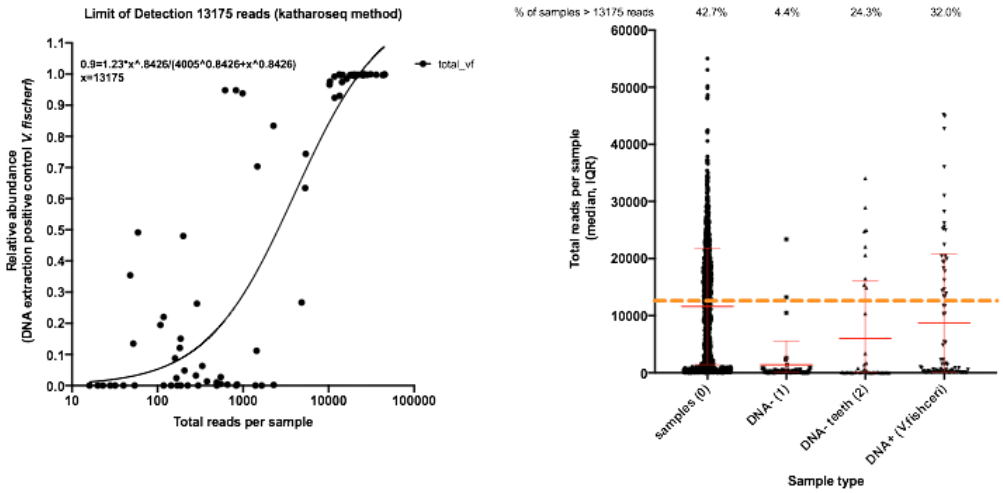


Figure 33 . Sample success from 16S microbiome analysis across sample types. (a) Katharoseq protocol applied to 10 DNA extraction plates determines that the sample cutoff criteria is 13175 reads. Thus samples with less than 13175 reads are excluded from the analysis. (b) 42.7% of tooth samples had at least 13175 reads and thus were included in the analysis. Read count distributions across sample types as related to sample cutoff read count (13175 reads)

.

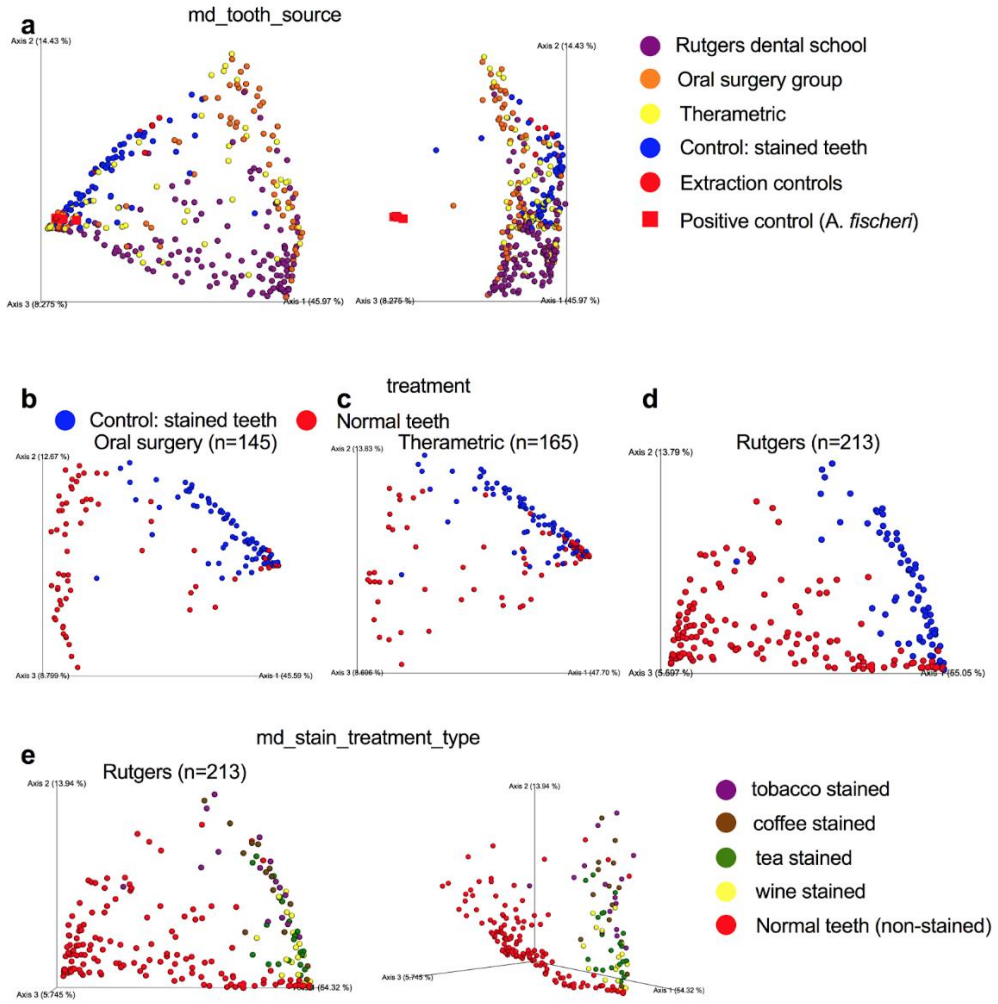


Figure 34. Primary drivers of teeth microbiome. a) Source of tooth acquisition is a primary driver of microbiome (ANOISM R2 0.331, P=0.001) followed by treatment (ANOISM 0.135, P=0.001). Independent analyses of each tooth source was conducted. Statistical testing of effects of stained vs. normal teeth for each tooth source (b) Oral Surgery Group (ANOISM R2=0.310, P=0.001), (c) Therametric (ANOISM R2=0.089, P=0.001), and (d) Rutgers dental school (ANOISM R2=0.434, P=0.001). (e) The normal non-stained tooth samples from Rutgers dental school were also significantly different than the combined stained teeth samples (ANOISM R2=0.261869, P=0.001).

Tooth source location is a major driver of teeth, thus analyses should be independent. Stained teeth are very different than unstained teeth across all sampling sites Rutgers, Oral surgery group, and Therametric.This suggests that staining alone does not account for the microbial diversity found on the teeth from the primary 'normal' samples (samples with missing measurements across teeth stain are removed). Stained teeth have a different microbial profile and specifically a gradient of differences across wine, tea, coffee, and tobacco.

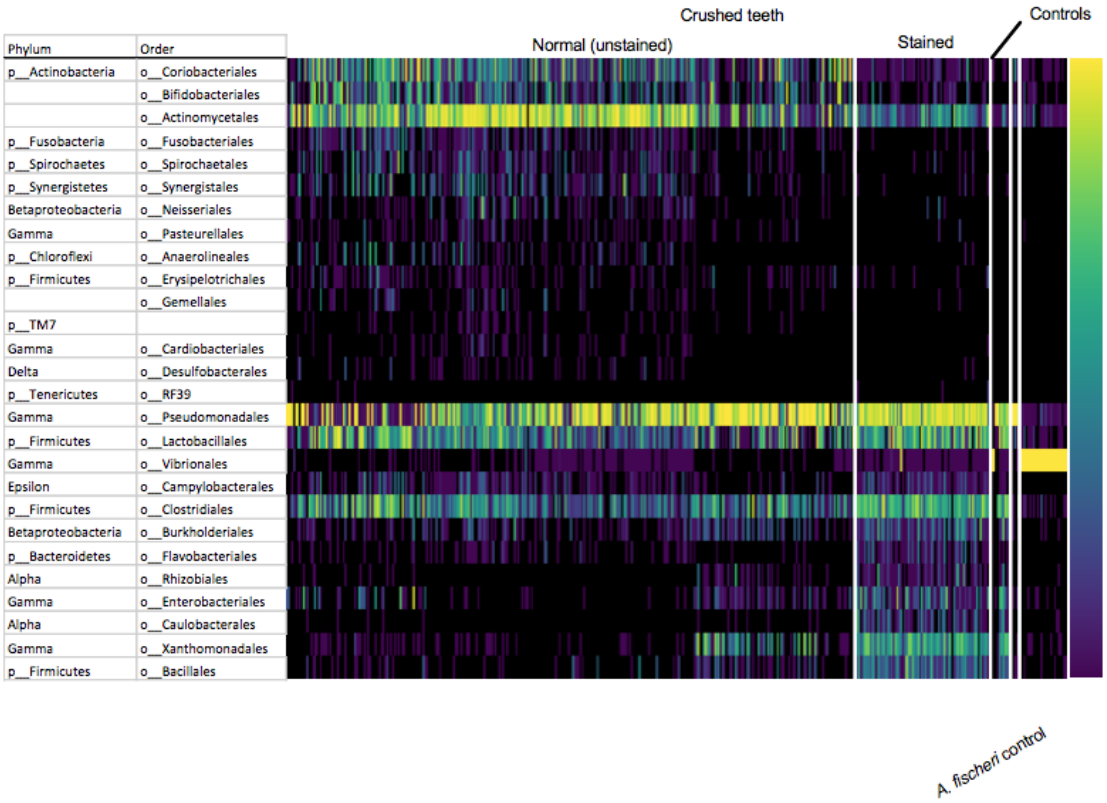


Figure 35. Order Map Tooth samples from all three sources (Therametric, Rutgers, and Oral surgery group) were analyzed to determine microbes which were associated with artificial staining procedures (tea, wine, tobacco, coffee). Differentially abundant sOTUs associated with artificial staining procedures were identified using Calour. The sOTUs were then collapsed down to the order level to show primary phylogenetic groupings of these microbes in context to controls in the study.

Overall teeth were most dominated by Pseudomonadales, Actinomycetales, Clostridiales, Lactobacillales. Normal (non-stained teeth) teeth had higher proportions of Actinobacteria and several other groups. Artificially stained teeth had higher proportions of Pseudomonadales, Lactobacillales, and several other orders.

Table 5. Sample Group Statistical Significance

The Statistical analysis of primary factors (md_teeth_source > staining method > if its stained) points at tooth source (md_teeth_source) as the most significant factor. The analyses of staining methods on independent sampling sources shows the Oral Surgery group (e.g. therametric, oral, rutgers) as the least significant contributor. All samples within each collection cohort were significantly different than stained teeth.

Table 1. Multivariate statistical analysis of metadata categories driving the tooth microbiome						
					Generalized_Unifrac	
Sample-subset	Sample_size	Metadata_category	type	test_method	R2	P-value
All samples	n=371	md_teeth_source	categorical	ADONIS	0.229	0.001
		md_stain_treatment_type	categorical	ADONIS	0.115	0.001
		Treatment	categorical	ADONIS	0.095	0.001
		a_astrix	continuous	ADONIS	0.994	0.006
		b_astrix	continuous	ADONIS	0.994	0.009
		l_astrix	continuous	ADONIS	0.994	0.012
		shade_trend	continuous	ADONIS		0.125
		wio	continuous	ADONIS	0.994	0.006
Therametric	n=165	md_stain_treatment_type	categorical	ANOISM		0.633
		Treatment	categorical	ANOISM	0.089	0.001
	Normal (n=85)	a_astrix	continuous	ADONIS	1.000	0.002
		b_astrix	continuous	ADONIS	1.000	0.001
		l_astrix	continuous	ADONIS	1.000	0.001
		shade_trend	continuous	ADONIS	1.000	0.007
		wio	continuous	ADONIS	1.000	0.001
Oral surgery group	n=145	md_stain_treatment_type	categorical	ANOISM		0.548
		Treatment	categorical	ANOISM	0.310	0.001
	Normal (n=73)	a_astrix	continuous	ADONIS		0.847
		b_astrix	continuous	ADONIS		0.828
		l_astrix	continuous	ADONIS		0.842
		shade_trend	continuous	ADONIS		0.836
		wio	continuous	ADONIS		0.826
Rutgers dental school	n=213	md_stain_treatment_type	categorical	ANOISM	0.262	0.001
		Treatment	categorical	ANOISM	0.434	0.001
	Normal (141)	a_astrix	continuous	ADONIS	1.000	0.001
		b_astrix	continuous	ADONIS	1.000	0.001
		l_astrix	continuous	ADONIS	1.000	0.001
		shade_trend	continuous	ADONIS	1.000	0.001
		wio	continuous	ADONIS	1.000	0.001

Therametric and Rutgers collected teeth had significant differences across the various staining measures (indicating microbial associations with staining), which is probably due to a sample size effect.

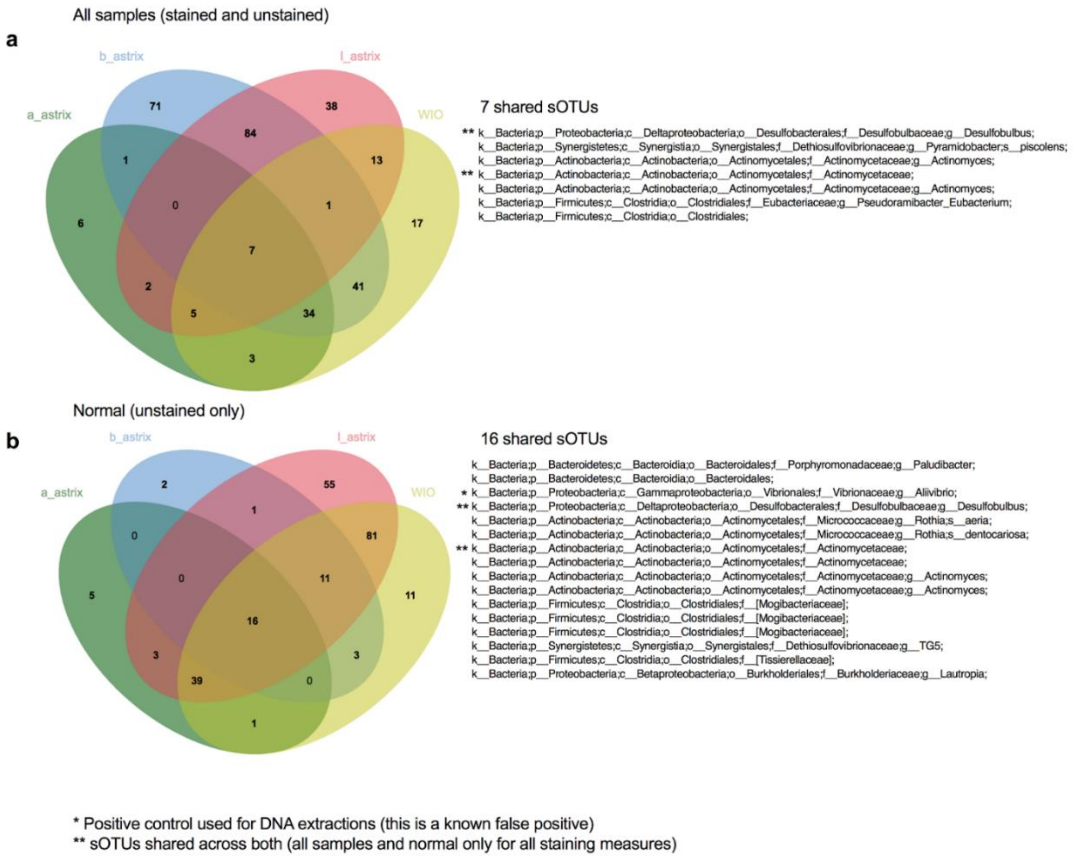


Figure 36. Microbial sOTUs associated with staining measures in artificially stained and normal teeth. Spearman correlations of sOTUs to various staining measures (a*, b*, l*, and wio). (a) Results from shared sOTU correlates when including all samples (stained and unstained) from the three locations Therametric, Oral surgery group, and Rutgers dental school. (b) Shared correlates across the staining measures when only performing correlations on the treatment 'Normal' samples (excludes all stained teeth). * indicates a sample that was used as a positive control in DNA extraction and is not a true positive signal. ** indicates sOTUs which were present as core correlates in both analyses (all samples and stained only).

This plot organizes the microbial correlates to demonstrate how many correlates are also shared across the various staining measures. When including all samples, 7 sOTUs were correlated with all four staining measures while 16 sOTUs were correlated with all 4 staining measures when using Normal (unstained only). Note, (shade_trend) did not yield any sOTU correlations.

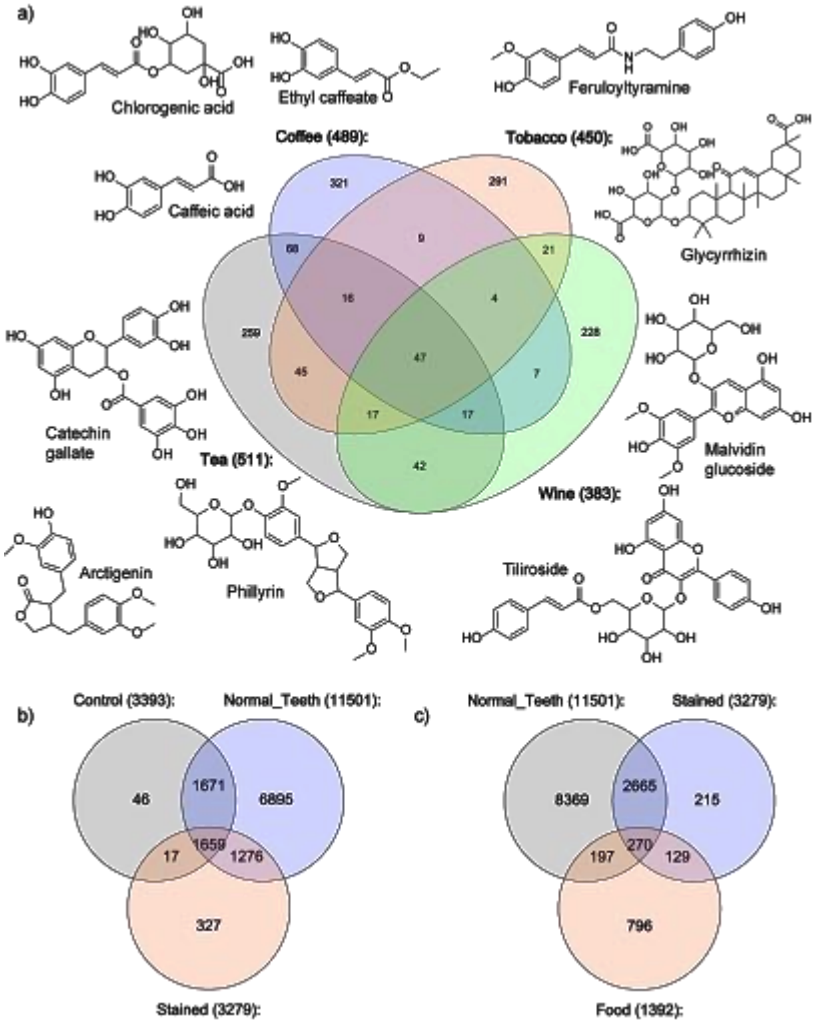


Figure 37. Venn diagrams of molecular networking results. a) Staining samples and molecules found uniquely in each staining solution. b) The overlap between regular, stained teeth, and hydroxyapatite powder. c) The overlap between normal, stained teeth, and staining solutions (food).

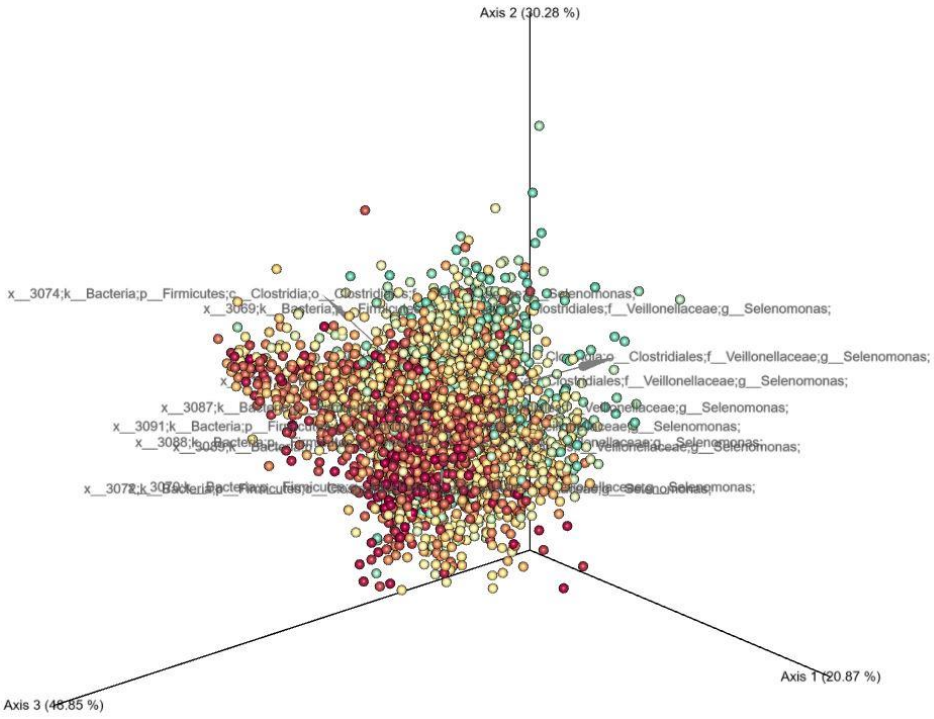


Figure 38. Selenamonas within the A Scale (Rutgers)
Different *Selenamonas* genus can be found within the ends of the staining scale for green and red. The microbial groups are found pointing towards Axis 1 and away from Axis 3. The A gradient trends away from Axis 2.
**Green is represented by the darker red while the green represents red stains.*

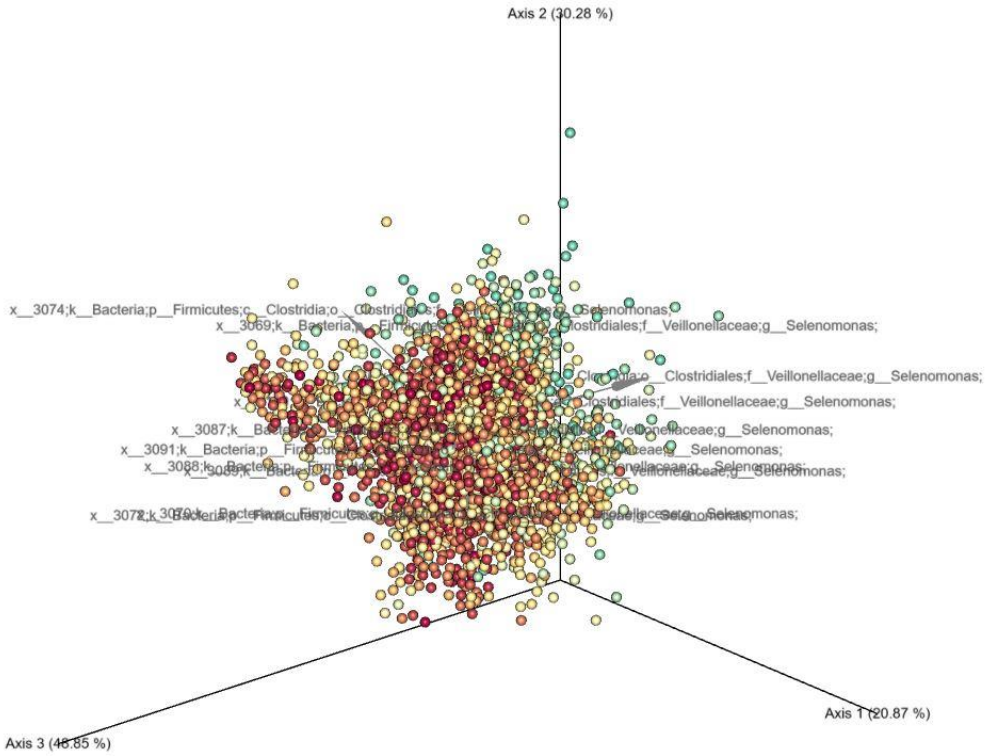


Figure 39. Selenamonas within the B Scale (Rutgers)
Different *Selenamonas* genus can be found within the ends of the staining scale for yellow and blue. The microbial groups are found pointing towards Axis 1 and away from Axis 3. The B gradient trends away from Axis 2.
**Yellow is represented by the darker red while the green represents blue stains.*

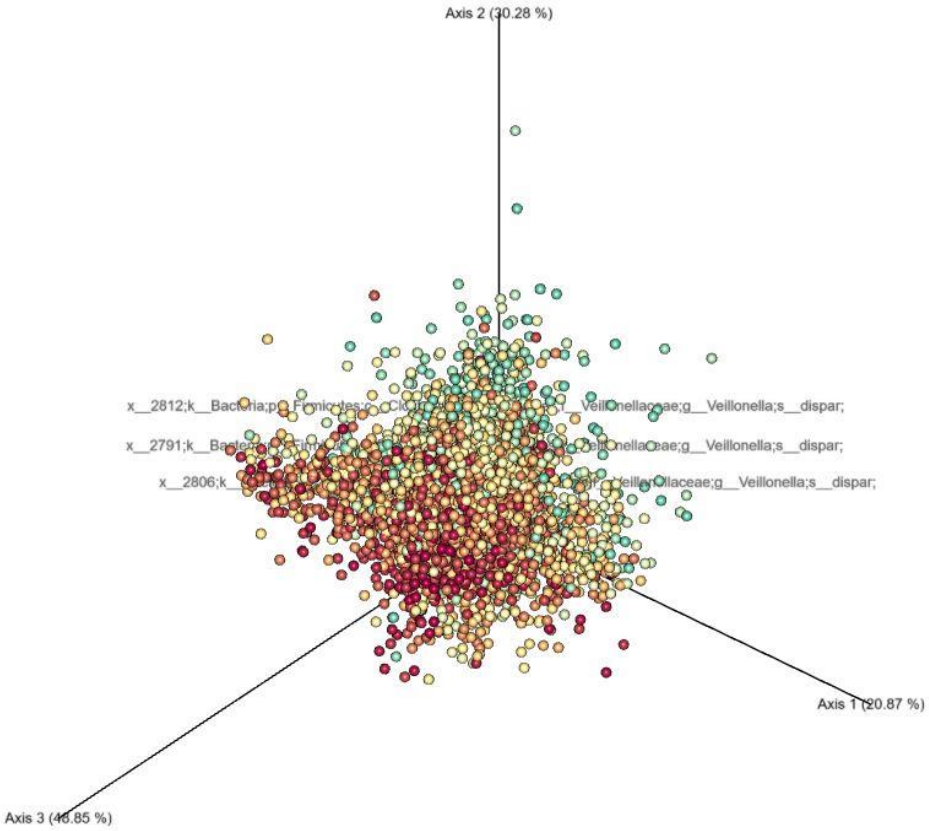


Figure 40. *Veillonella Dispar* within the A scale (Rutgers).
Dispar tend to be found closer to the green points which correlate with green stains. Microbial arrows are pointing away from Axis 3. The L gradient trends away from Axis 2.

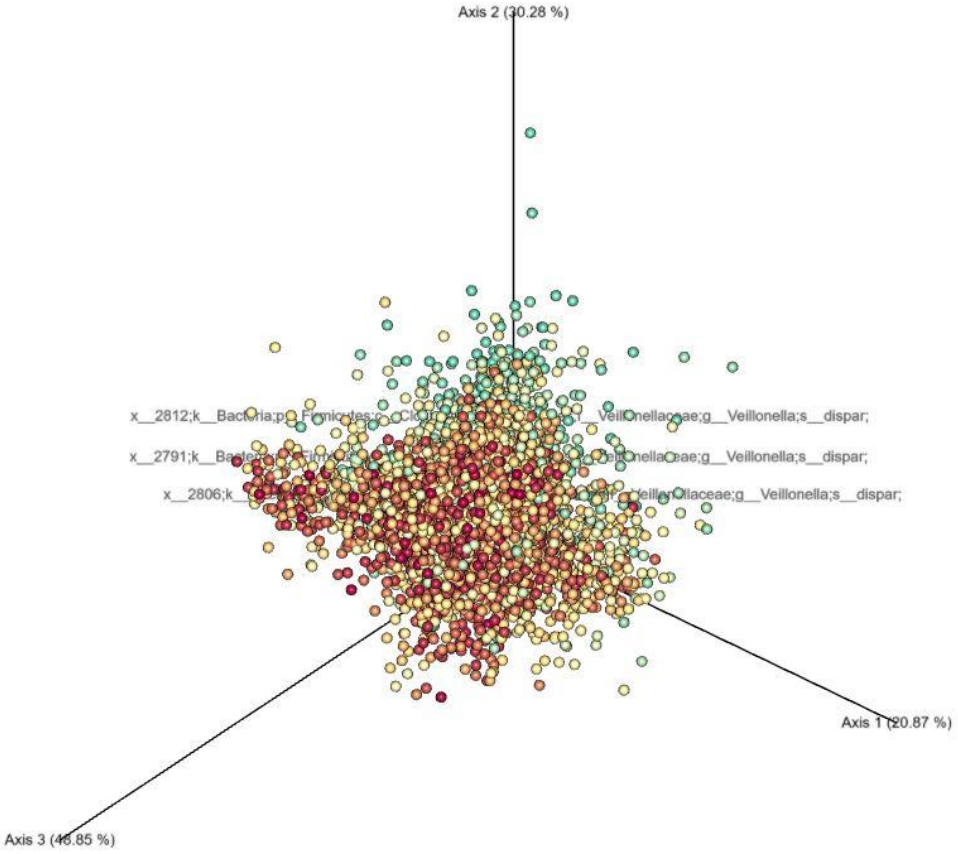


Figure 41. *Veillonella Dispar* within the B scale (Rutgers).
Dispar tend to be found closer to the green points which correlate with blue stains. Microbial arrows are pointing away from Axis 3. The B gradient trends away from Axis 2.

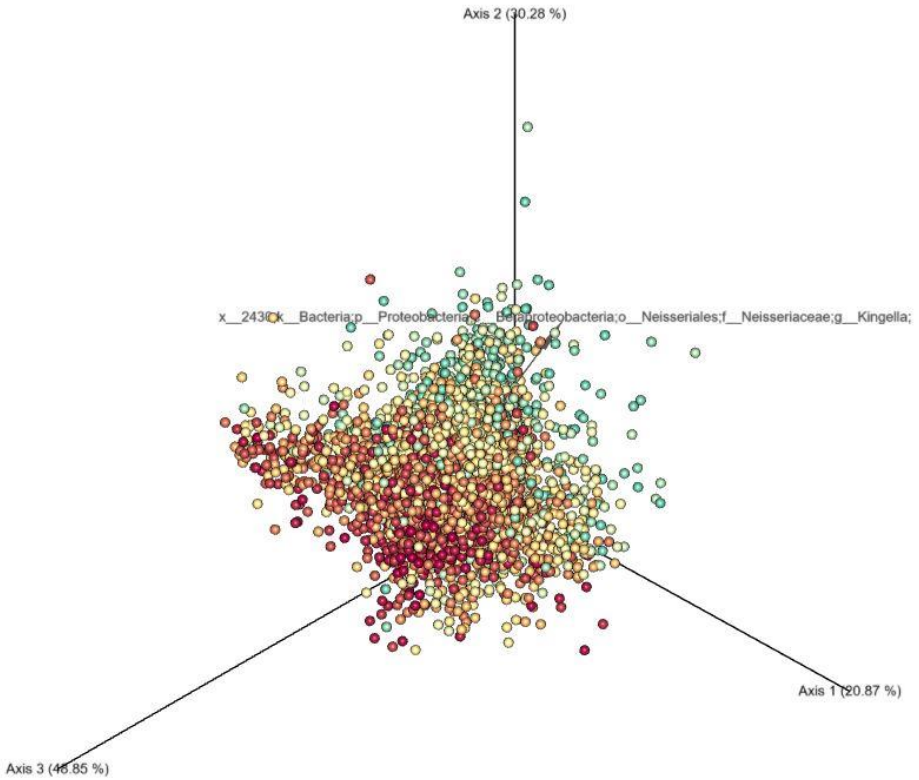


Figure 42. *Kingella* within the A scale (Rutgers).
The *Kingella* genus tends to be found closer to the green points which correlate with red stains. The microbial arrow is pointed toward Axis 1. The A gradient trends away from Axis 2.

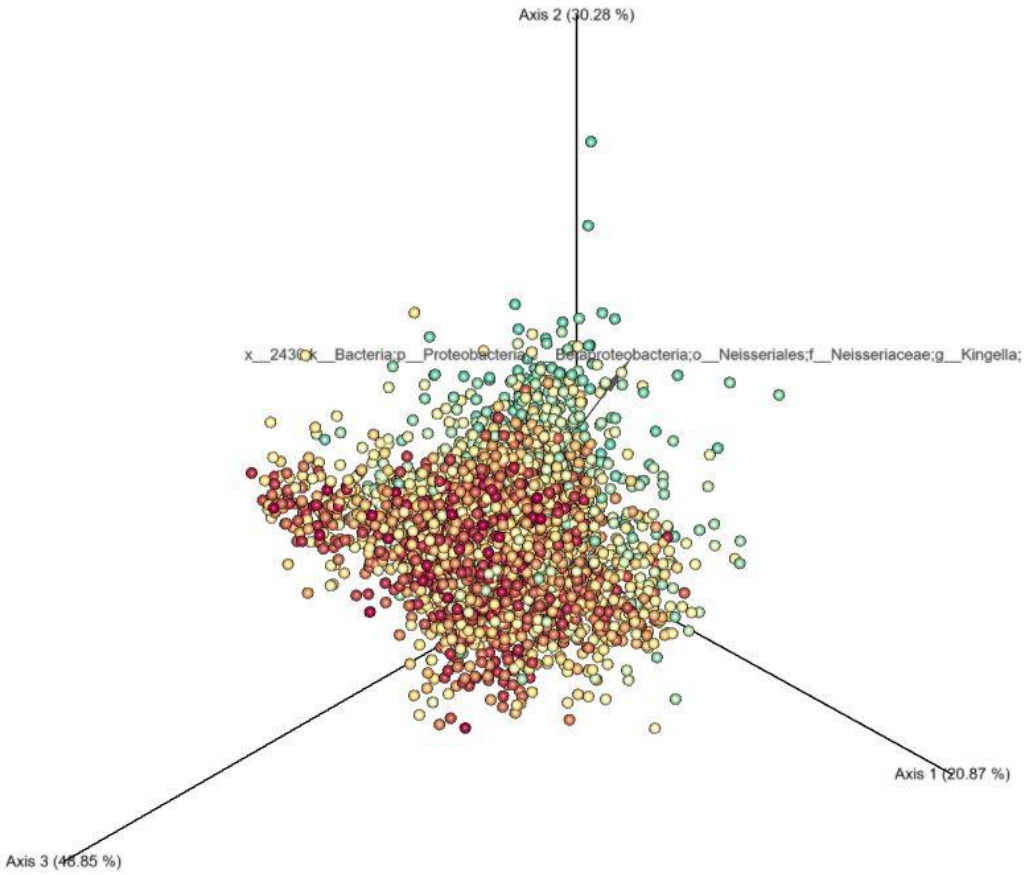


Figure 43. *Kingella* within the B scale (Rutgers).
The *Kingella* genus tends to be found closer to the green points which correlate with blue stains. The microbial arrow is pointed toward Axis 1. The B gradient trends away from Axis 2.

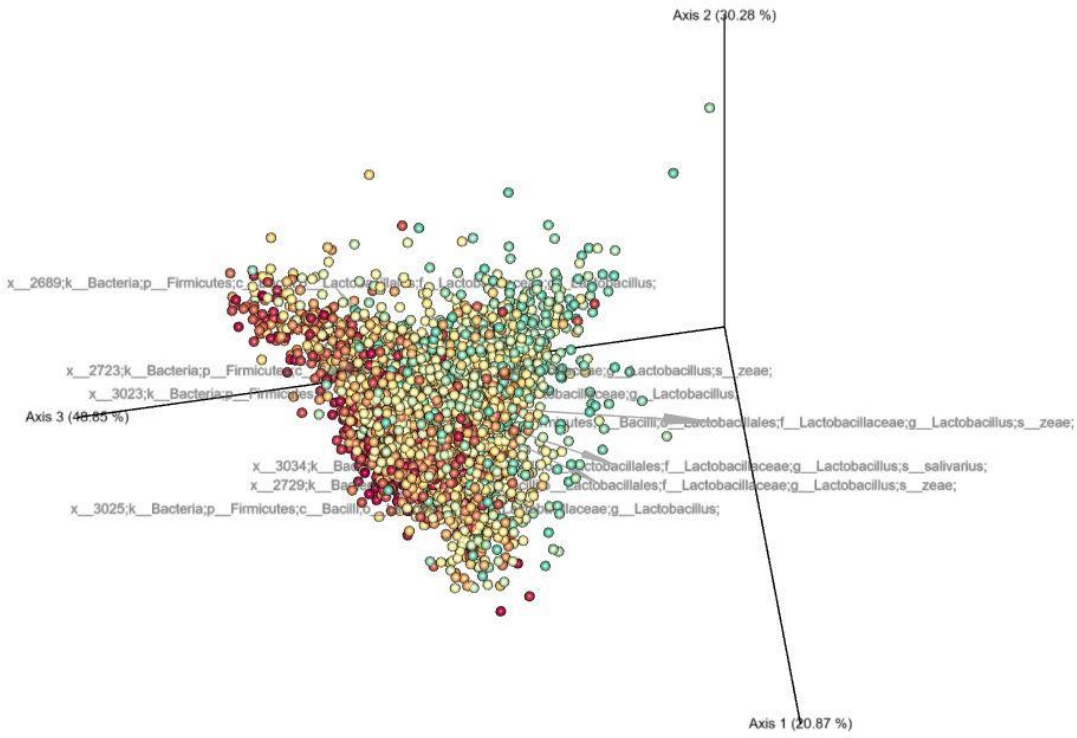


Figure 44. *Lactobacillus zeae and salvarius* within the A scale (Rutgers). *Zeae and Salvarius* tend to be found closer to the green points which correlate with green stains. The microbial arrows point away from axis 3 in the same direction as the green points. The A gradient trends away from Axis 2.

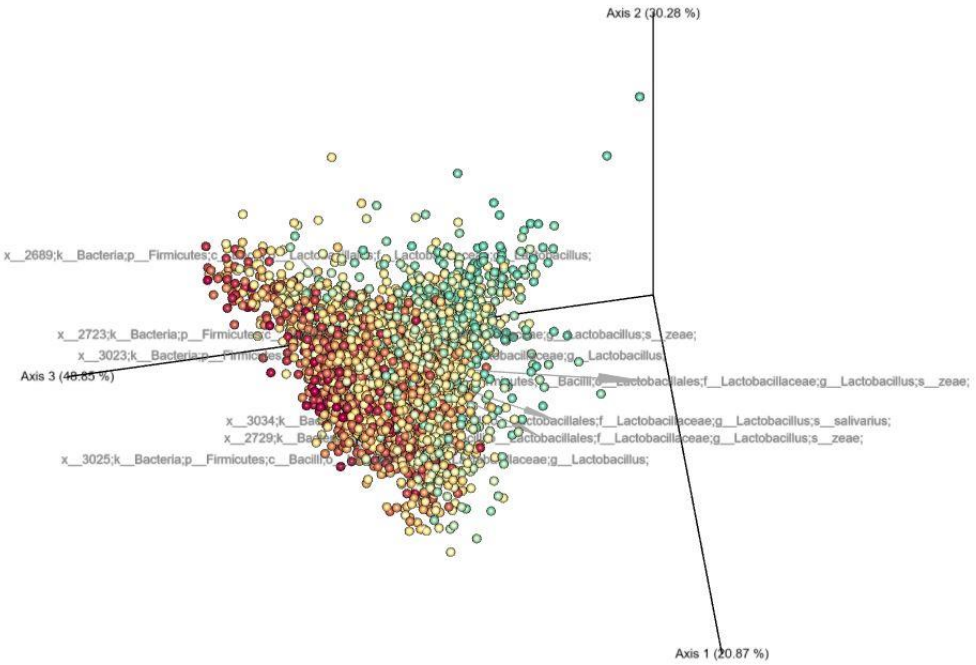


Figure 45. *Lactobacillus zeae and salvarius* within the B scale (Rutgers). *Zeae and Salvarius* tend to be found closer to the green points which correlate with blue stains. The microbial arrows point away from axis 3 in the same direction as the green points. The B gradient trends away from Axis 2.

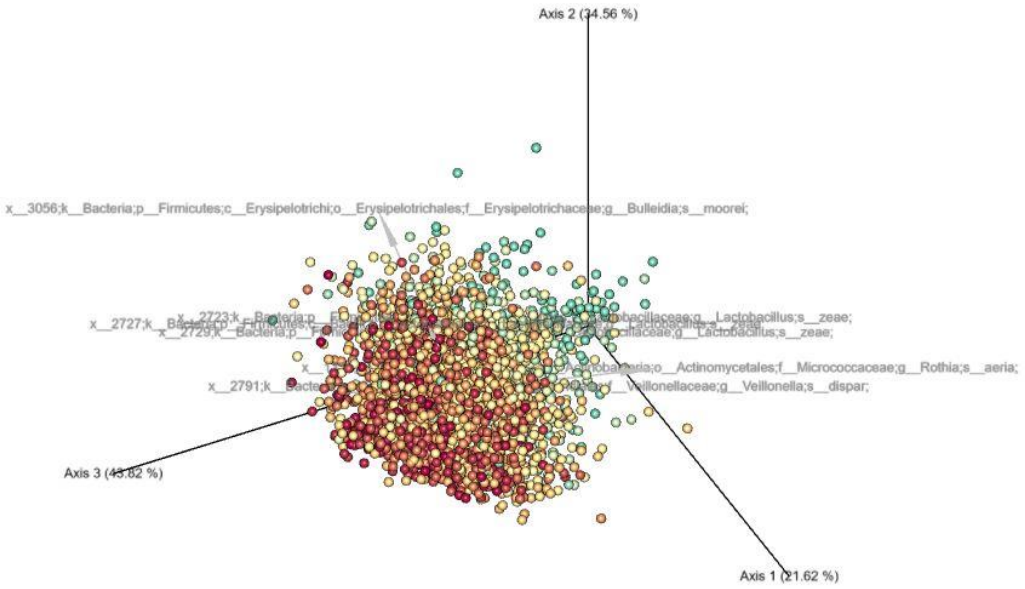


Figure 46. Species within the Therametric dataset on the A scale.
Lactobacillus zeae, Rothia aeria, and Bulleidia moorei (Solobacterium moorei) all showed to be significant for the metabolites contributing to staining. *Moorei* and *Zeae* have the most influence ongreen staining while *Aeria* has an impact on red staining.

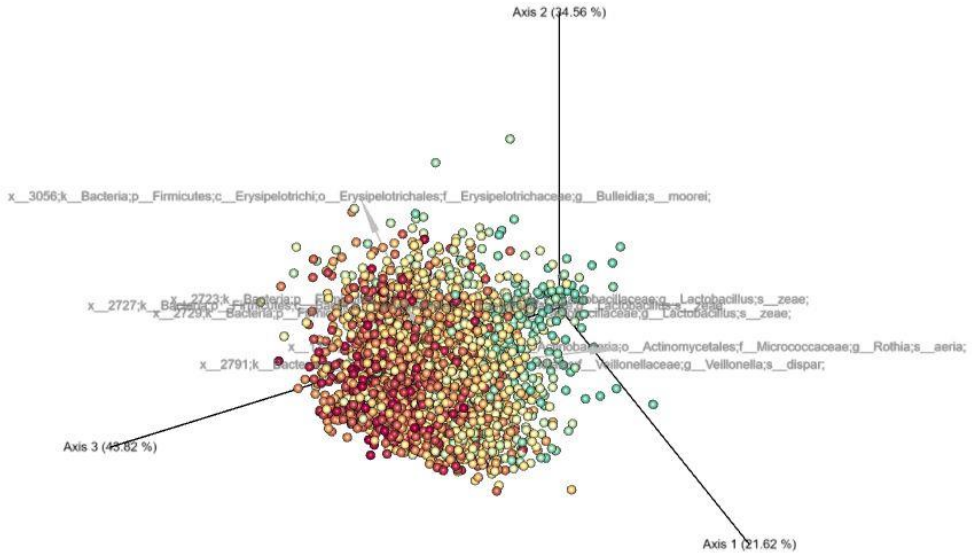


Figure 47. Species within the Therametric dataset on the B scale.
Lactobacillus zeae, Rothia aeria, and Bulleidia moorei (Solobacterium moorei) all showed to be significant for the metabolites contributing to staining. *Moorei* and *Zeae* have the most influence on yellow staining while *Aeria* has an impact on blue staining.

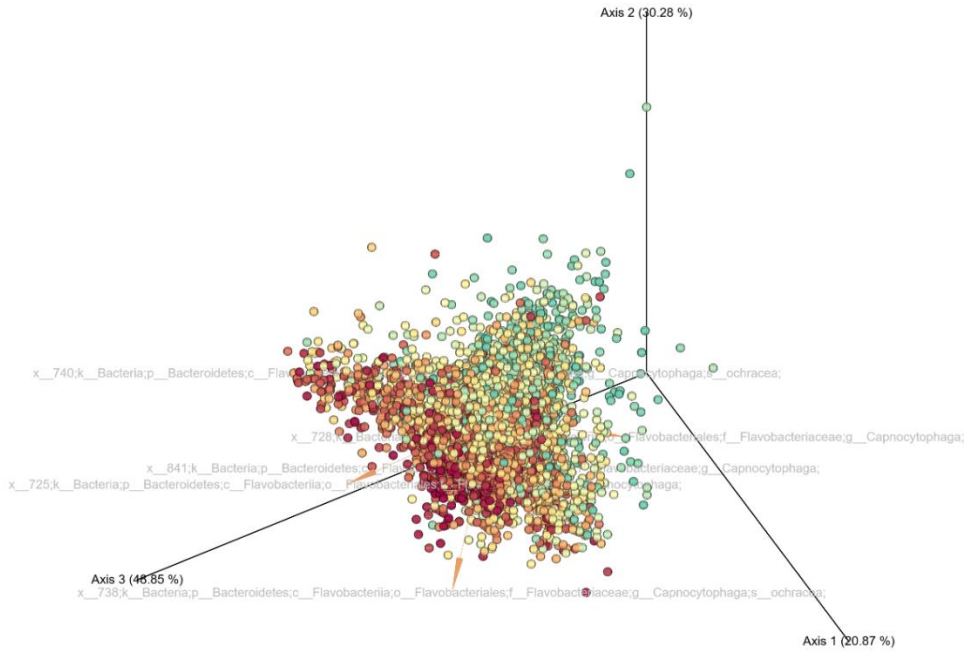


Figure 48. *Capnocytophaga* and *C. Ochracea* within the A scale (Rutgers) *Capnocytophaga* and *C. Ochracea* significantly correlated for the metabolites contributing to staining. Both highly contributed to green staining and show orange arrows pointed away from Axis 3. The A gradient is going towards Axis 2 from green staining to red staining.

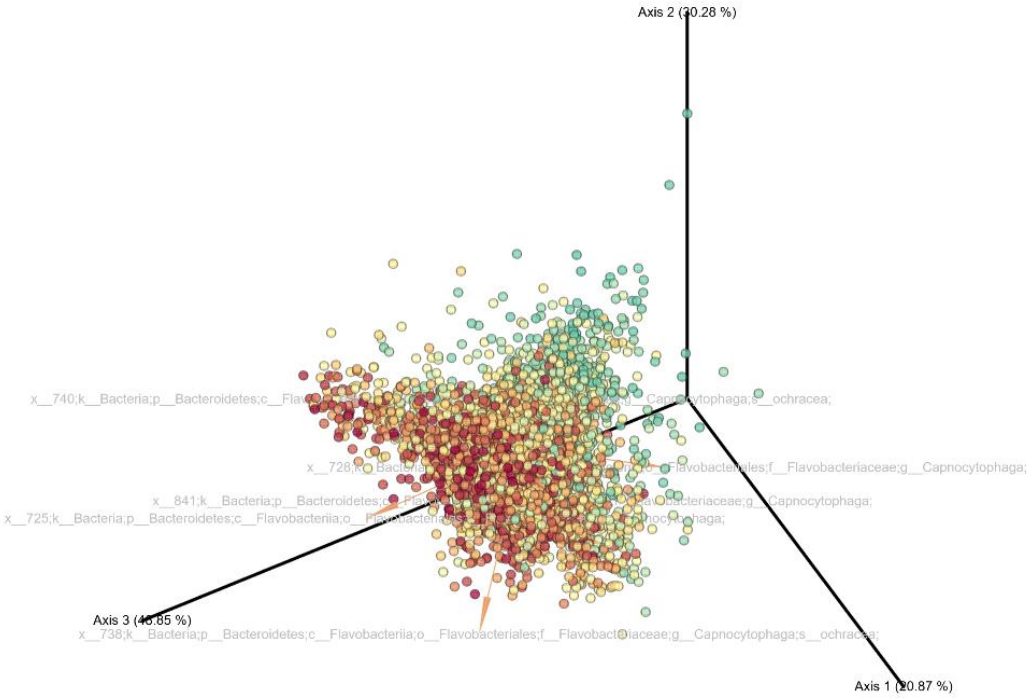


Figure 49. *Capnocytophaga* and *C. Ochracea* within the B scale (Rutgers)*Capnocytophaga* and *C. Ochracea* significantly correlated for the metabolites contributing to staining. Both highly contributed to yellow staining and show orange arrows pointed away from Axis 3. The B gradient is going towards Axis 2 from yellow staining to blue.

Figure 50. *Eikenella* within the A scale (Rutgers)
The Eikenella spp. significantly correlated for the metabolites contributing to staining. It highly contributes to green staining and show a green arrow pointed away from Axis 3. The A gradient is going towards Axis 2 from green staining to red.

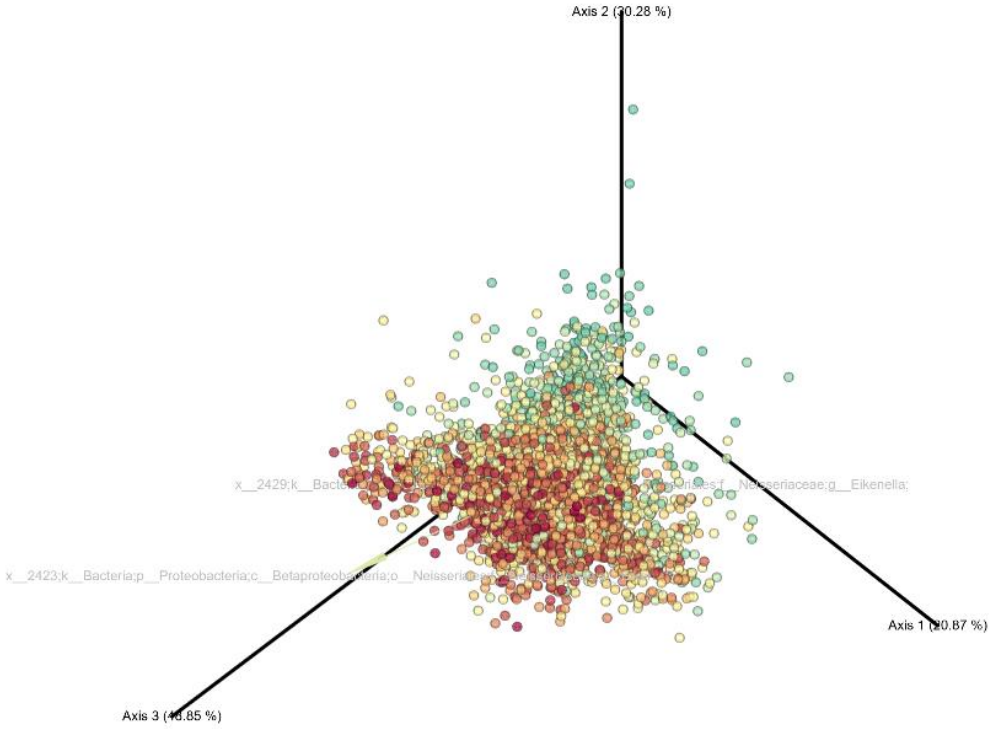


Figure 51. *Eikenella* within the B scale (Rutgers)
The Eikenella spp. significantly correlated for the metabolites contributing to staining. It highly contributes to yellow staining and show a green arrow pointed away from Axis 3. The B gradient is going towards Axis 2 from yellow staining to blue.

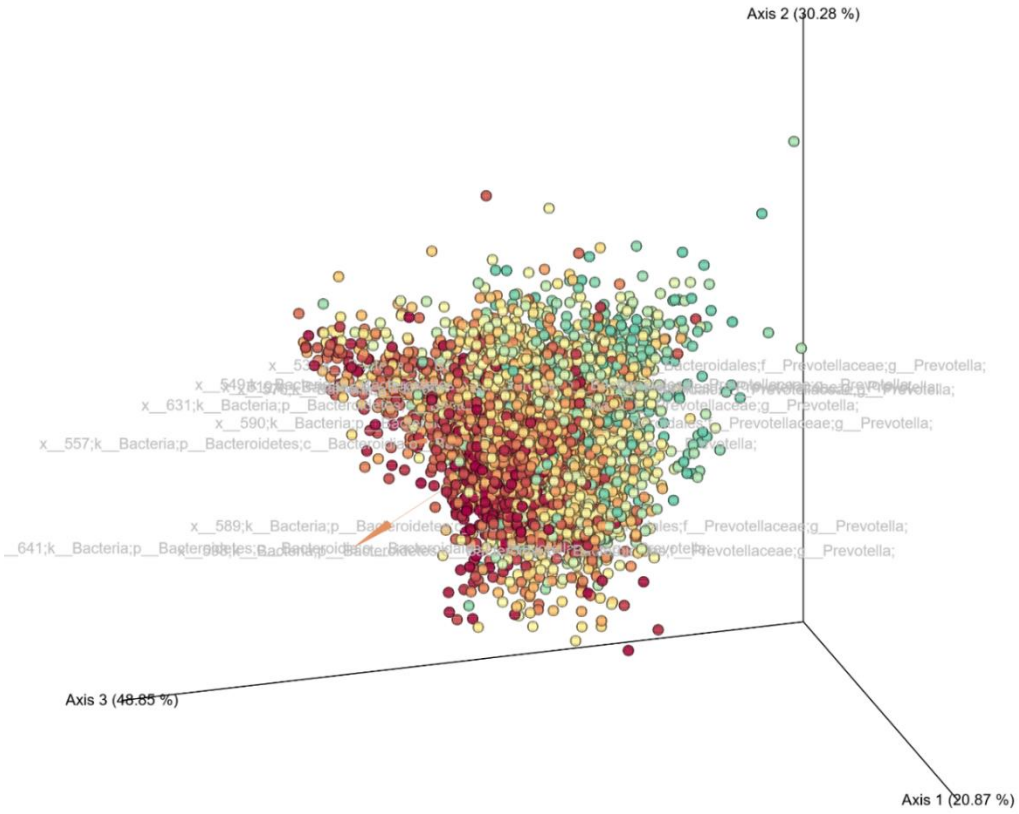


Figure 52. *Prevotella spp.* within the A scale (Rutgers)
The Prevotella spp. significantly correlated for the metabolites contributing to staining. It highly contributes to green staining and shows an orange arrow pointed away from Axis 3. The A gradient is going towards Axis 2 from green staining to red.

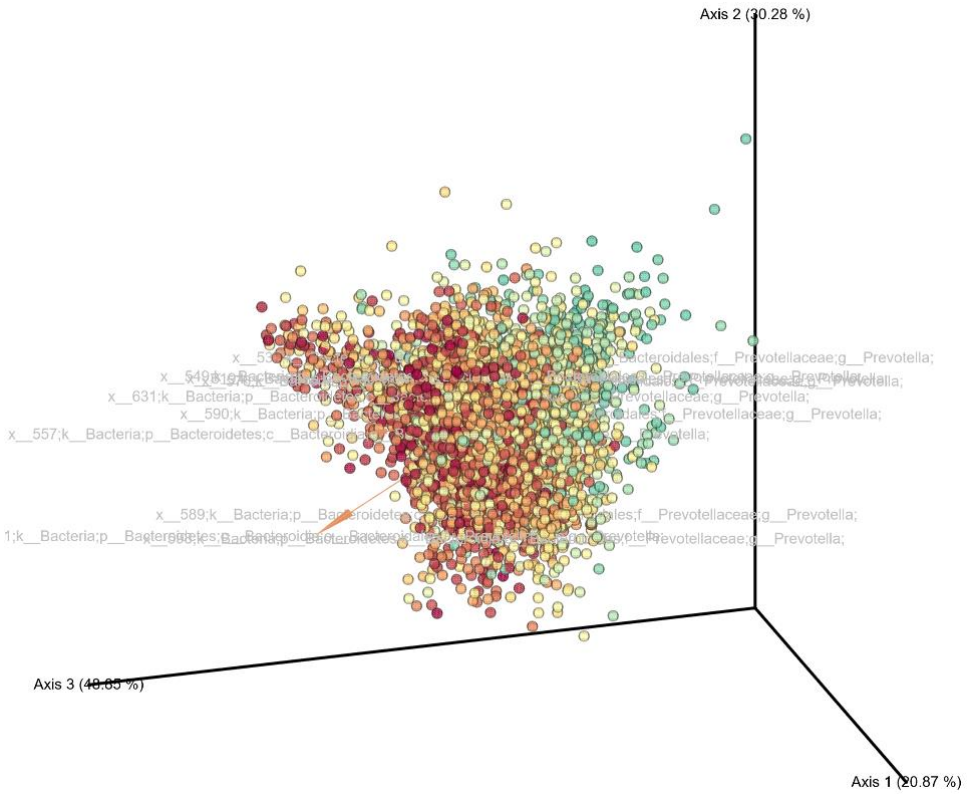


Figure 53. *Prevotella spp.* within the B scale (Rutgers)
The Prevotella spp. significantly correlated for the metabolites contributing to staining. It highly contributes to yellow staining and shows an orange arrow pointed away from Axis 3. The B gradient is going towards Axis 2 from yellow staining to blue.

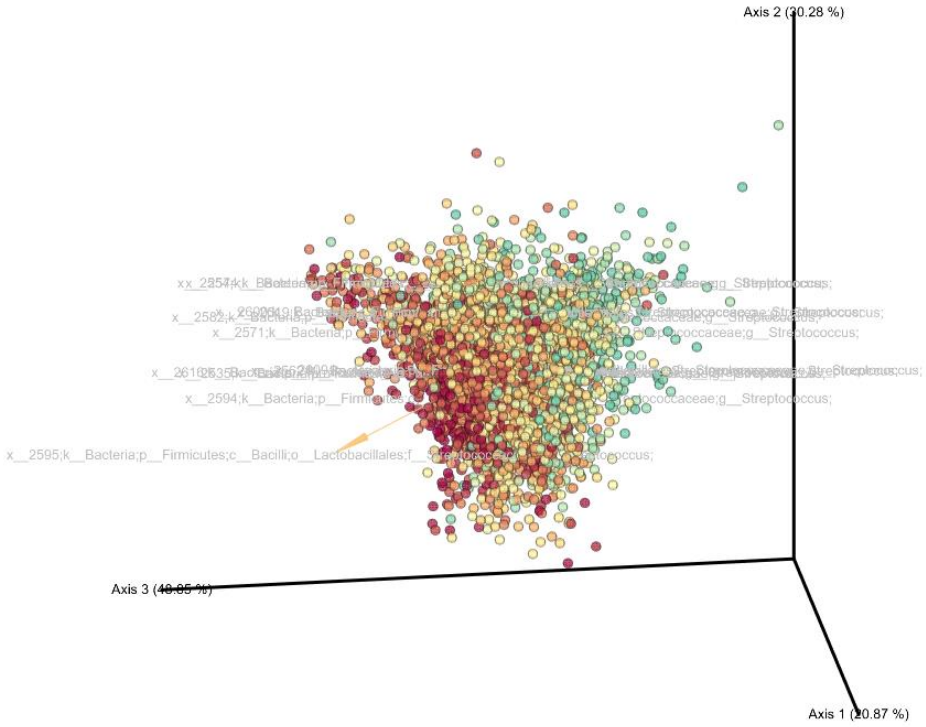


Figure 54. *Streptococcus spp.* within the A scale (Rutgers)
The Streptococcus spp. significantly correlated for the metabolites contributing to staining. It highly contributes to green staining and shows an orange arrow pointed away from Axis 3. The L gradient is going towards Axis 2 from green staining to red.

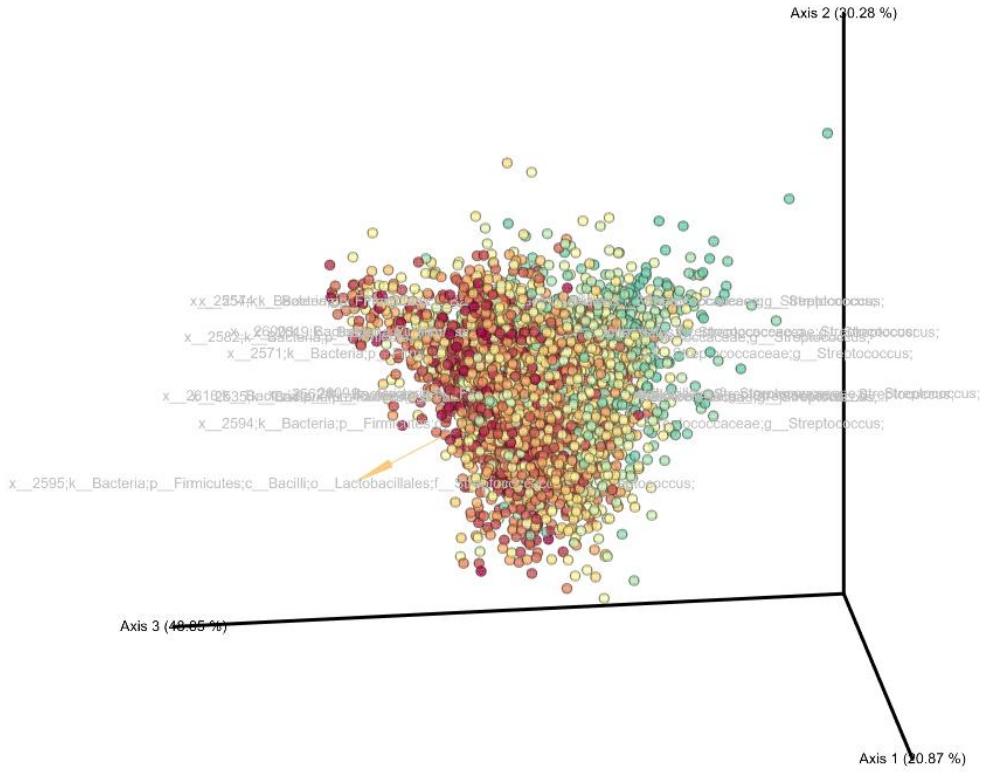


Figure 55. *Streptococcus spp.* within the B scale (Rutgers)
The Streptococcus spp. significantly correlated for the metabolites contributing to staining. It highly contributes to yellow staining and shows an orange arrow pointed away from Axis 3. The B gradient is going towards Axis 2 from yellow staining to blue.