RAPID IDENTIFICATION OF *SALMONELLA* SEROTYPES USING FTIR AND ANN TECHNOLOGY

A University Thesis Presented to the Faculty

of

California State University, East Bay

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Biological Science

By

Bijan Kashanian

June 2017

RAPID IDENTIFICATION OF *SALMONELLA* SEROTYPES

RAPID IDENTIFICATION OF *SALMONELLA* SEROTYPES

USING FTIR AND ANN TECHNOLOGY

By

Bijan Kashanian

Approved:	Date:
Carol Lauzon	5/19/17
Claudia Uhde-Stone	2017 May 19
Lisa Gorski	5-19-2017

ACKNOWLEDGEMENTS

First and foremost, I wish to thank **Dr. Carol Lauzon** for providing me the opportunity to pursue this project. Without her unconditional support and ability to inspire, I certainly would not have made so much progress academically or professionally. Thank you, **Dr. Lisa Gorski**, for allowing the collaboration which this project would not exist without. I would also like to thank **Dr. Claudia Uhde-Stone** for joining my committee and working with me to write this thesis. A very special thanks to **Dr. Shunlin Wang** of Bruker Optics for making something as seemingly complicated as Artificial Neural Network analysis of Fourier Transform Infrared Spectroscopy easily understandable and incredibly intuitive.

The United States Department of Agriculture – ARS Laboratories (Albany, CA) Gorski Lab was invaluable in this project. Their welcoming, friendly, and safe lab environment was the foundation that supported the development of my thesis work. To my mentor, **Kelly Romanolo**, who was pivotal to my success, I couldn't have done this without you. Thank you for teaching me many of the techniques for this project as well as helping troubleshoot any issues that arose. I can't forget **Samar Walker** and **Anita Liang** who maintained an unwavering positivity and fostered moral support throughout the many hours spent processing samples; the lab would have been a much lonelier place without their friendship. Thank you so much!

Thank you, **Ashley Maynard**, for keeping me motivated through dozens of caffeine supplemented all-nighters. There were so many moments where I felt like giving up, but she never let me lose hope that results were just around the corner.

To my **Mom**, thank you for just, well, being my mom. I'm sure it wasn't easy trying to keep me out of trouble, you deserve a medal for getting us through everything. To my **Dad**, thank you for raising me with the notion that education is a virtue. I doubt I'll ever want to stop learning. Words can't describe the gratitude I have for both of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 BACKGROUND

Salmonella is a genus of rod-shaped, Gram-negative, flagellated, facultatively anaerobic bacteria (Baron, 1996). The genus is composed of two major species: *enterica* and *bongori*. Of these species, *enterica*, is further divided into six subtypes based on differential characteristics: *enterica* (serotype I), *salamae* (serotype II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI) (Grimont & Weill, 2007). *Salmonella enterica* subtypes are further divided by serotypes based on serological variants or serovars. There are currently over 2500 different *Salmonella* serovars, 1547 of which belong to *Salmonella enterica* subsp. *enterica* (Guibourdenche et al., 2010).

Salmonella enterica is one of the most commonly occurring food-borne pathogens in the United States (Scallan, 2011). A multitude of *Salmonella enterica* serotypes contribute to an estimated 1.2 million illnesses each year in the United States (CDC, 2011). Detection and identification of *Salmonella enterica* as well as maintenance and improvement of sanitary conditions are imperative to quell a *Salmonella* outbreak (Bertrand et al., 2010). In humans, the top twenty most prevalent *Salmonella* serotypes contributed to 49,000 lab-confirmed infections in the United States (CDC, 2012). The disparity in estimated infections versus lab-confirmed infections could be due to the broad range of infection symptoms that lead to many unreported cases.

Pathogenesis & Virulence Factors

In order for *Salmonella* spp. to begin colonization of the human host, they must first enter the body through ingestion of contaminated food or from person-to-person

contact (Baron, 1996). Upon ingestion, *Salmonella* spp. must overcome a gauntlet of host defenses in order to colonize the human gut (Hallstrom & McCormick, 2011). A thick mucosal layer covers the intestinal epithelium, physically impeding invading microorganisms from epithelial cell engagement while also harboring antimicrobial substances. In addition to the challenge of traversing the mucosal layer, the colonizing pathogens must compete with the existing microbiota for nutritional resources as well as attachment sites to the intestinal epithelium (Ahmer & Gunn, 2011).

While the battery of initial host defenses appears to be extremely limiting for invasive internalization of pathogens, *Salmonella* spp. possess pathogenicity islands (a distinct class of genomic islands acquired through horizontal gene transfer) that serve as molecular utility boxes for overcoming and evading these resistances (Groisman & Ochman, 1996). One major factor toward overcoming gut colonization resistance used by pathogens is the induction of the inflammatory response in the host. The host inflammatory response can be triggered by effector molecules secreted from Type III Secretion Systems (T3SS) found on pathogenicity islands 1 and 2. Also, the detection of Pathogen- associated Molecular Patterns (PAMPS), such as lipopolysaccharide, peptidoglycan, and flagellin, found on the surface of invading *Salmonella* spp. (Abrahams & Hensel, 2006) influence the host inflammatory response. This induced inflammation lowers the population of the existing microbiota, which interfere with colonization resistance (Lupp et al., 2007). Conversely, resident gut microbiota are able to effectively out-compete *Salmonella* that lack inflammatory capabilities (Barman et al., 2008).

Following colonization of the mucosal surface of the gut, *Salmonella* will begin its next step of pathogenesis: internalization into host cells. *Salmonella* enters host cells through two distinct processes. Phagocytic cells can efficiently recognize and internalize *Salmonella*, or alternatively, *Salmonella* can utilize two distinct types of T3SS as a tool to invade phagocytic and non-phagocytic cells. T3SS-1 is utilized extracellularly while T3SS-2 is induced intracellularly (Ibarra et al., 2010). T3SS-1 is able to translocate at least 15 effector molecules into the host cell, some of which cooperatively induce actin rearrangements to promote rapid internalization (Ibarra & Steele-Mortimer, 2009). Following internalization, *Salmonella* remain in a modified phagosome, known as a *Salmonella*-containing vacuole (SCV), which resembles an early endosome. Within 60-90 minutes, these SCVs begin to mature and become highly enriched with markers of late endosomes and lysosomes (Steele-Mortimer et al., 1999).

Maturing SCVs possess the features of late endosomes, but that is the extent of their originally intended endosomal activity. SCVs modify their compartments with pathogenicity island derived proteins, arresting the host endosomal pathway at the late endosomal stage (Garai et al., 2012). As an SCV matures, it traffics itself along microtubules towards the microtubule-organizing center at the juxtanuclear position. The SCV provides a hospitable environment as well as nutrients for the *Salmonella* to replicate. Deiwick et al. (2006) hypothesized that the SCV obtains nutrients through fusion with other vesicles, as it requires an intact golgi apparatus to elicit pathogen replication.

Typhoidal and Non-typhoidal *Salmonella*

Salmonella enterica serovars can be categorized by the diseases and distinct immune responses they elicit and are grouped as either typhoidal or non-typhoidal. *Salmonella enterica* serovars, such as Typhi, Sendai, and Paratyphi A, B, or C (collectively referred to as typhoidal serovars) are specialist pathogens that use human hosts as their exclusive reservoir. Typhoidal infections cause enteric fever, an invasive and life-threatening systemic disease. Developing regions that lack clean water and adequate sanitation are often particularly susceptible to typhoidal *Salmonella* outbreaks. In contrast, non-typhoidal *Salmonella* (NTS) serovars are generalist pathogens with broad host specificity and occur worldwide (Gal-Mor et al., 2014).

Typhoidal *Salmonella* can persist in a host's body, causing chronic infection for even decades after infection. The gallbladder is a suspected site of persistence for *S.* Typhi due to production of gallstones. Gallstones also provide a site for pathogen attachment and are conducive to the formation of biofilms. Biofilms protect microorganisms from immune responses and antimicrobial compounds, and thus develop a carrier state (Gonzalez-Escobedo et al., 2011).

Symptoms of NTS infections can range from mild abdominal pains to severe dehydration caused by diarrheal-based water loss. NTS infections may require hospitalization, and can sometimes become severe enough to cause bacteremia (Sirinavin et al., 1999). While most *Salmonella* infections originate from ingestion of contaminated food, person-to-person spread is also possible. Symptoms of salmonellosis generally occur 6 to 48 hours after ingestion of *Salmonella* spp. and most often take the form of nausea, vomiting, diarrhea, and abdominal pain; less common symptoms include fevers

and chills. While some specific serotypes cause specific symptoms, it is possible for any serotype to produce any of the possible symptoms (Baron, 1996).

Serotyping

Identification of *Salmonella* strains can be a daunting task. Serotyping by slide agglutination is a phenotypic method of subtyping that utilizes surface antigens that react with specific antibodies to cause the respective molecules to adhere and aggregate. Serotyping involves characterization of somatic (O) and flagellar (H) antigens that can differ among each strain of *Salmonella*. The use of serotyping over a large time span allows for long-term epidemiological surveillance of *Salmonella* outbreaks (Wattiau et al., 2011), however, using this method requires complete O and H antigen characterization for final identification, resulting in the use of numerous resources and prolonged procedure times (Difco Laboratories, 1998). In fringe cases, "inconsistent" variants often require complementary investigations to address unexpected expression levels of surface antigens (Schrader et al., 2008 and Barco et al., 2014). Thus, the serotyping method, while useful, can be expensive due to the cost of reagents and antibodies, and been found to sometimes lead to inconclusive results.

Pulse-Field Gel Electrophoresis (PFGE) is a molecular typing method considered as the current "gold standard" for identifying *Salmonella* (Wattiau et al., 2011). This method utilizes banding patterns of genomic DNA obtained from cells grown to an optimal concentration followed by cell lysis, treatment with a specific restriction enzyme or enzymes, then made to run on a PFGE lane (CDC, 2013). The most common restriction enzyme selected to digest the genomic DNA is XbaI, which has a probability

of cutting every 4,100 base pairs. When analyzed through traditional agarose gel electrophoresis, DNA fragments are separated by size via an electrical field, but an indistinguishable banding pattern is generated due to overlapping large fragments.

PFGE allows for separation of large fragments by utilizing an alternating voltage gradient to improve resolution (Schwartz & Cantor, 1984). Through PFGE, specific serotypes will contain fragments of distinguishable sizes, forming a unique banding pattern that serves as a fingerprint for that serotype. Banding patterns specific to strains or serotypes are highly distinctive in most cases, leading to successful tracking of outbreaks (Threlfall et al., 1999). While slide agglutination can be costly, PFGE is relatively inexpensive to perform and widely reproducible. However, PFGE is labor-intensive and time consuming. Furthermore, despite PFGE being the gold standard, serotype discrimination of a few *Salmonella* serotypes is too low to draw firm conclusions (Kérouanton et al., 2007).

Despite their drawbacks, these methods are some of the most common and reliable tools for serotype characterization. If a superior alternative were to emerge, it would need to produce faster results that rival, if not surpass, the accuracy of the previously mentioned methods. This alternative should also be cheaper to utilize, require less labor, and have higher throughput than previous methods. Ideally, all results would be easily reproducible and easy to distribute among other laboratories. A strong candidate for an alternate method of identification and discrimination is Fourier transform infrared spectroscopy (FTIR).

FTIR utilizes the differential excitation of molecules or compounds based on the range of frequencies of infrared light used to excite them. Any given molecule will

absorb different frequencies of infrared light based on their chemical structure; these absorptions are resonant frequencies. Simple, non-Fourier Transform absorption spectroscopy utilizes a monochromatic light source to measure how much light is absorbed by the molecule in question; absorption measurements must be taken for each desired wavelength separately. FTIR spectroscopy allows acquisition of absorption with the spectrum from 0 to 4000 cm^{-1} simultaneously (Koenig, 1981). With FTIR technology, one can obtain unique spectra of the different biochemical components and functional groups of intact cells and utilize this spectral profile to characterize the cells in question (Kim et al., 2005).

While FTIR can produce hundreds of spectra in a short amount of time, thus reducing wet-lab preparation time, this efficiency highlights a new bottleneck in bacterial subtyping. Analysis of acquired spectra, if done manually, could take a tremendous amount of time. However, an emerging solution comes in the form of Artificial Neural Networks (ANN). ANN is a relatively modern computational tool that can be applied to many disciplines in order to model and process large data sets, and even solve complex problems (Basheer & Hajmeer, 2000). Utilizing FTIR to produce spectra and ANN to analyze them, the bottleneck in this process becomes sample acquisition and preparation time as opposed to the analysis of results.

1.2 STATEMENT OF THESIS

The potential for FTIR to be utilized as a tool for rapid identification and serotyping is an alluring prospect with a high likelihood of utilizing online collaboration as a method of constructing and growing a global library for microbiological

classification. FTIR can identify and discern S. *aureus* from other species of *Staphylococcus* (Lamprell et al., 2006), and has been utilized with ANN to rapidly serotype *Liseria* spp. (Romanolo et al., 2015).

The two objectives of my research were to:

(1) Explore the use of FTIR as an accurate method of obtaining spectra from 18 *Salmonella* serovars selected from a list of the "Top 20" most prevalent strains responsible for food-borne Salmonellosis (Table 1, modified from CDC, 2012) and from 24 replicates of *Salmonella enterica enterica* Kentucky and,

(2) Determine the capacity of an Artificial Neural Network to reliably learn, recognize, and classify the spectra for each corresponding serovar.

Table 1: The twenty most frequently reported Salmonella serotypes in the US in 2011, modified from CDC Salmonella annual report (2012). Serotypes denoted with an asterisk were used in this study.

Rank	Serotype	Number Reported	Percent
1	Enteritidis*	7553	16.5
2	Typhimurium*	6131	13.4
3	Newport*	5211	11.4
4	Javiana*	2937	6.4
5	4, [5], 12:i:-*	1339	2.9
6	Montevideo*	1196	2.6
7	Heidelberg*	1103	2.4
8	Muenchen*	984	2.1
9	Infantis*	910	2
10	Branderup*	739	1.6
11	Oranienburg*	721	1.6
12	Saintpaul*	709	1.5
13	Mississippi	549	1.5
14	Thompson*	536	1.2
15	Agona*	505	1.1
16	Paratyphi B var. L(+) tartrate+*	431	0.9
17	Bareilly	429	0.9
18	Typhi	393	0.8
19	Berta*	321	0.7
20	Anatum*	282	0.6
	Sub Total	**32969**	**71.9**
	All Other Serotyped	6864	15
	Unknown	4173	9.1
	Partially serotyped	1603	3.5
	Rough, mucoid, and/or nonmotile isolates	219	0.5
	Sub Total	**12859**	**28.1**
	Total	**45828**	**100**

2 MATERIALS AND METHODS

2.1 *SALMONELLA ENTERITIDIS* CULTURE PREPARATION

All *Salmonella* strains were provided by Dr. Lisa Gorski and were identified and serotyped previously, and isolated at the USDA Microbiology Research Unit in Albany, CA. Strains were stored in 1 mL cryotubes (USA Scientific, Ocala, Florida) containing 100 uL glycerol and 900 uL of bacterial isolate at -80°C. Each trial contained up to 24 biological replicates (different strains) of a serovar (Table 2). Each replicate was gently thawed before being aseptically applied to a Tryptic Soy Agar (TSA, Beckton-Dickinson-BBL, Franklin Lakes, NJ) plate and incubated overnight (18-24 hours) at 37°C. Isolates of each serotype in log phase were selected for analysis. One loopful of each *Salmonella* culture was taken aseptically from a colony on a TSA plate and inserted into a 10 mL tube containing 7mL of sterile Tryptic Soy Broth (TSB). Each culture was incubated for 18 hours at 37°C, shaking at 150 RPM in a Multitron Pro shaking incubator (Infors, Bottmingen, Switzerland).

Table 2: Trials organized by date. Each trial contained at least one serovar with 24 biological replicates and 72 technical replicates. Trials 14, 15, and 19 contained more than one serovar.

Trial	Date	Serovar	Biological Replicates	Technical Replicates
1	5-Feb-15	Enteritidis	24	72
2	12-Feb-15	Typhimurium	24	72
3	19-Feb-15	Newport	24	72
4	26-Feb-15	Montevideo	24	72
5	5-Mar-15	Heidelberg	24	72
6	12-Mar-15	Montevideo	24	72
7	19-Mar-15	Muenchen	18	54
8	26-Mar-15	Javiana	6	18
9	2-Apr-15	Infantis	24	72
10	9-Apr-15	Oranienburg	24	72
11	16-Apr-15	Saintpaul	24	72
12	23-Apr-15	Thompson	24	72
13	30-Apr-15	Anatum	24	72
14	7-May-15	Agona	20	60
14	7-May-15	4,5,12:i:-	3	9
15	14-May-15	Braenderup	13	39
15	14-May-15	4,5,12:i:-	10	30
16	21-May-15	Enteritidis	24	72
17	28-May-15	Typhimurium	24	72
18	4-Jun-15	Newport	24	72
19	11-Jun-15	Berta	12	36
19	11-Jun-15	Javiana	6	18
19	11-Jun-15	Paratyphi B	6	18
20	18-Jun-15	Kentucky	24	72
21	25-Jun-15	Enteritidis	24	72
	Total	**19**	**478**	**1434**

2.2 SAMPLE CLEAN UP

After 18 hours of incubation, 1.5 mL of each replicate were transferred into wells within a 2 mL deep 96-well plate (Eppendorf AG, Hamburg, Germany) and covered with an aluminum seal (Excel Scientific, Inc. Victorville, CA). The 96-well plate was centrifuged at 3000x RCF for 20 minutes with a 5804 Benchtop Centrifuge (Eppendorf AG, Hamburg, Germany). The aluminum seal was subsequently removed in order to access the supernatant. Supernatant from each well was removed as thoroughly as possible without disturbing the pellet formed through centrifugation and discarded. One mL of 0.85% saline solution was added to the wells to re-suspend the pellet. The plate was then covered with another aluminum seal and centrifuged for a second time at 3000x RCF for 20 minutes. The process was repeated an additional two times. After the final centrifugation, the supernatant was carefully removed and the pellet re-suspended with 100 uL double distilled H_20. Contents were then transferred into a 300 uL 96-well ELISA plate (Corning Incorporated, Corning, NY). Optical density readings were taken for each sample using a BioTek Synergy HT Microplate Reader (Biotek, Winooski, VT) at 630 nm. For sample consistency, a standard optical density between 1.5-2.0 was obtained for each sample.

2.3 FOURIER-TRANSFORM INFRARED SPECTROSCOPY ANALYSIS

Five uL of each sample were plated in triplicate within a 384-well silicon microplate (Bruker Optics, Ettlingen, Germany). Seventy-two samples were loaded, in total, for most serotypes. The microplate was then placed in an incubator to dry for approximately 30 minutes. The dried microplate was then loaded into the Bruker FT-IR

Spectrometer's HTX-XT high-throughput extension module (Bruker Optics) and analyzed using OPUS/LAB software (Bruker Optics).

2.4 ARTIFICIAL NEURAL NETWORK ANALYSIS

Spectra obtained through FTIR were used to train the Artificial Neural Network software, NeuroDeveloper™ (Syntheon, Heidelberg, Germany) for rapid identification of select *Salmonella* serotypes. NeuroDeveloper applies multiple processing techniques aimed to reduce the data points in each spectra from a vast amount down to a select number of unique points that will be used as features for differentiation. However, situations arise where two or more serotypes are too spectrophotometrically similar; using a larger set of features is often necessary to thoroughly mitigate this resemblance. First, a Savitzky-Golay smoothing and derivative filter is applied with 9 smoothing points to increase the signal-to-noise ratio of each spectra without greatly distorting the signal. Vector-normalization across the entire spectral range is followed by a covar selection process. Lastly, the spectral window is modified to exclude the ranges between 1800 cm^{-1} to 2800 cm^{-1}. This process reduces the spectra from 1261 points to 100 best suited points. The NeuroDeveloper™ software then pretests each input serotype for separation in order to determine if samples are suitable for further classification analysis. The pretest is performed in order to avoid wasting time and computational resources.

Initial pretests revealed that 100 points, and even 1000 unique points, were not enough for adequate separation. When serotypes are not separable, the software identifies which groups are causing the discrepancy. Five of the total serotypes (4,5,12:i:-, Javiana, Kentucky, Newport, and Paratyphi B) were identified as not separable and were

suggested to be grouped into a 2nd neural network. Repeating the same processing steps with the five serotypes in a separate network (Neural Net 2) allowed for a successful pretest for separation using 1000 data points.

After allowing the ANN to process each spectra and develop sufficient features, I was able to simulate automatic training. The training simulation consisted of three automation layers: (1) Input layer, which specifies how many points the software will draw conclusions and form algorithmic neurons through, (2) Hidden layer, which can be described as a connection point between the 'pre-synaptic' and 'post-synaptic' neurons that transforms the input signal into information which the output signal can use, and (3) Output layer, where the conclusion or response is drawn.

Neural Net 1 had an input layer set to 1000 neurons out of the 1000 available, a hidden layer consisting of 4 neurons with a logistic activation function, and an output layer consisting of 13 neurons with a logistic activation function. Neural Net 2 could only obtain 205 unique data points from the provided spectra, thus reducing the amount of input neurons. Neural Net 2 had an input layer consisting of 205 neurons of 205 possible, a hidden layer consisting of 13 neurons with a logistic activation function, and an output layer consisting of 5 neurons with a logistic activation function.

After the training session, I was able to test the ANN for its ability to assign serotypes based on their spectra. When classifying each spectra, the ANN analyzes the spectra with three mathematical algorithms to draw its conclusions: (1) WTA, (2) 402040, and (3) Extrapolation. When the spectra in question closely resembles a reference spectra, each algorithm will individually determine that they match, or pass the evaluation. The ANN uses a three-color rating system to denote the quality of the output to a correct

species using the three aforementioned mathematical criteria. Passing 3 out of 3 evaluations is denoted with a green dot, 2 out of 3 with a purple diamond, and 1 (or 0) out of 3 with a red triangle.

3 RESULTS

3.1 INFRARED SPECRAL ANALYSIS

I was able to obtain clean IR spectra for the following serotypes: Enteritidis, Newport, Typhimurium, Heidelberg, Montevideo, Javiana, and Muenchen. These trials were performed between the dates January 21st, 2015 and March 25th, 2015. All trials conducted following those dates appear to contain a large peak between 2200 cm^{-1} and 2400 cm^{-1} and all features appear more jagged and rough than the first seven (See Appendix A for examples sorted by date). A typical spectrum reflects the molecular components of bacterial cells with specific features corresponding to each components absorbance range (Figure 1).

All available isolates were successfully prepared and spectra successfully acquired. However, due to the limitations of some serotypes sample size, serotypes 4,5,12,:i:-, Berta, Braenderup, Javiana, and Paratyphi B had fewer isolates.

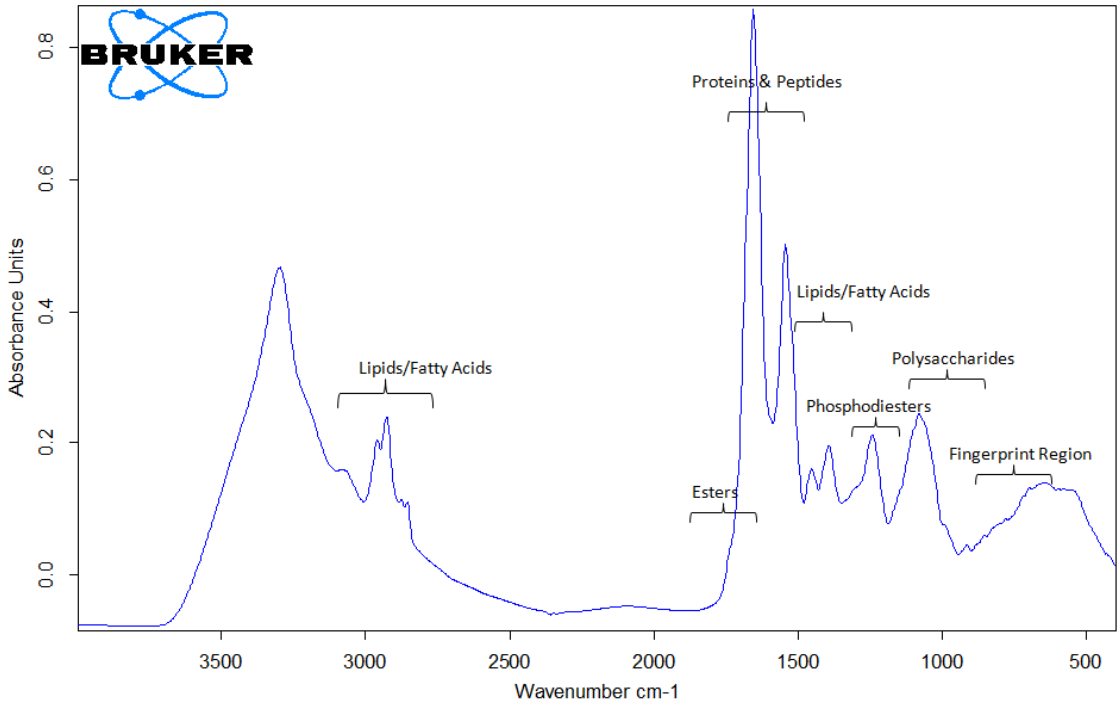


Figure 1: FTIR spectrum of Salmonella enteritidis. Regions corresponding to characteristic biomolecules are labeled.

3.2 ARTIFICIAL NEURAL NETWORK ANALYSIS

All spectral files were organized by trial date and serotype prior to loading into the ANN NeuroDeveloper™ software. Certain serotypes were not sufficiently separable, thus a second neural net was needed. All preprocessing steps were performed in the same order between the two neural nets in order to maintain consistency. Results of the preprocessing steps yield the first derivative of the original spectra (Figure 2), which allows the ANN to more easily determine 1000 unique points for each serotype classification.

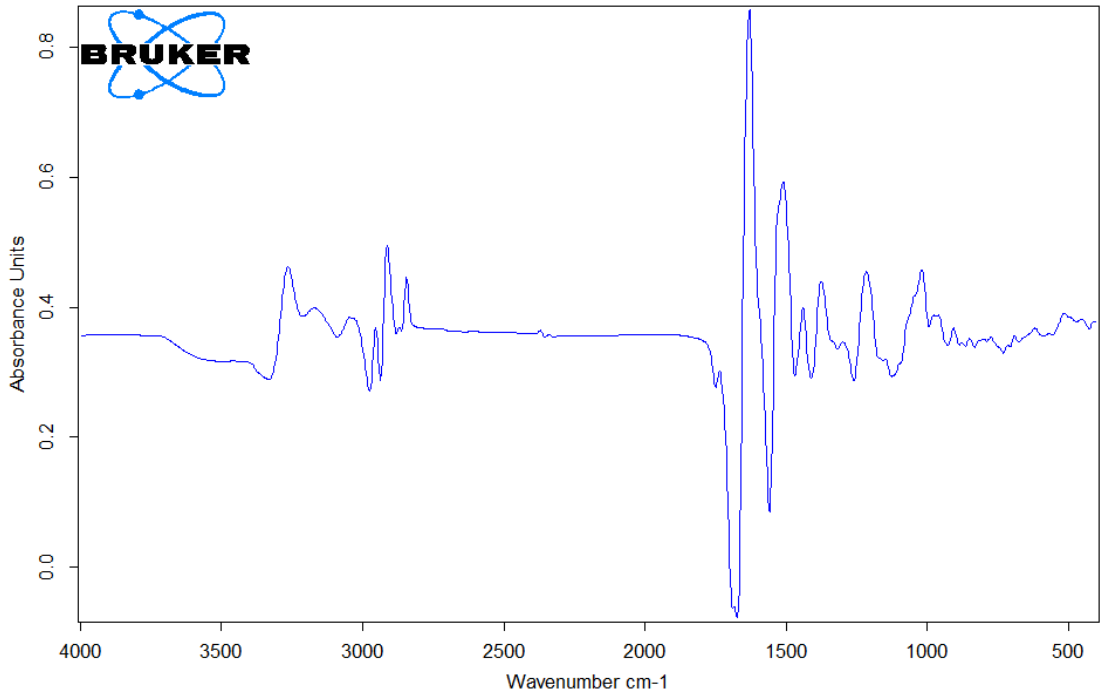


Figure 2: First derivative and smoothing processing of FTIR spectrum of Salmonella Enteritidis – original spectrum in Figure 1.

Of the 1038 spectra generated, 830 were assigned to classification using Neural Net 1 and 208 were assigned to Neural Net 2. Neural Net 1 contained serotypes that passed the initial preprocessing test for separation, and yielded a 97% total accuracy rate. Individual serotype classification accuracy shows the lowest accuracy belonging to *Salmonella* Braenderup while *Salmonella* Infantis, Montevideo, and Muenchen yielded 100% accuracy (Figure 3).

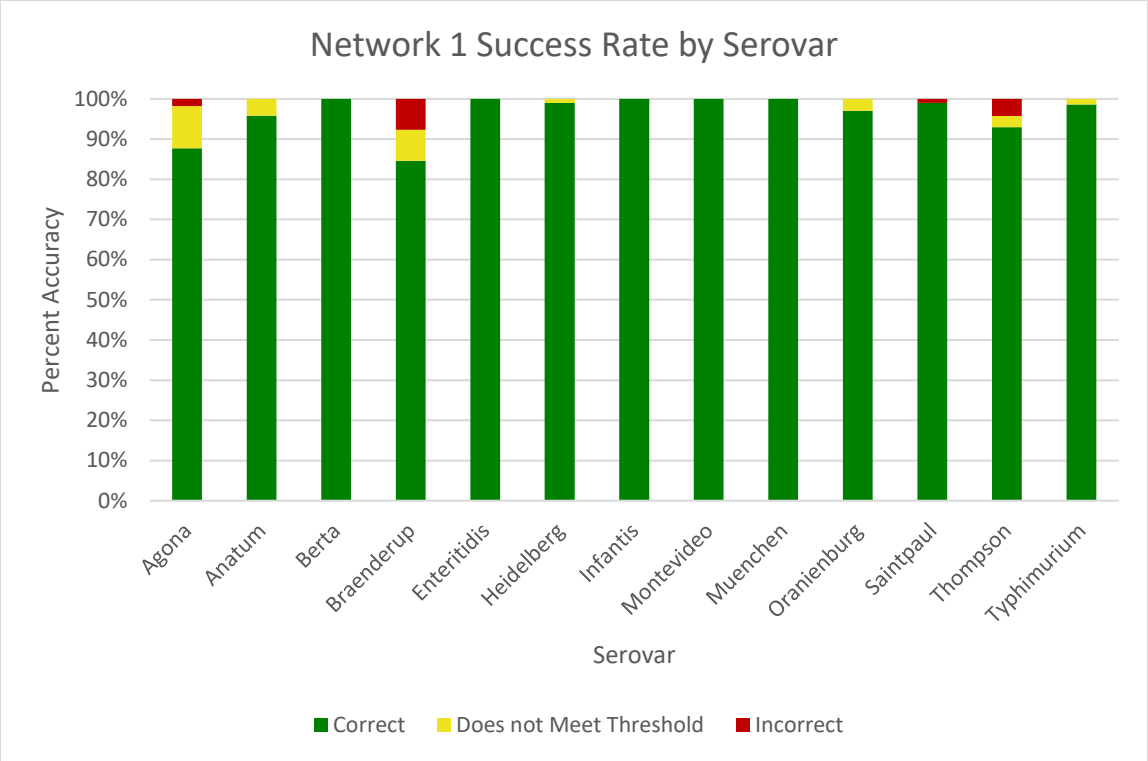


Figure 3: Success rate of Neural Net 1 organized by serovar. Green denotes correct identification; yellow denotes correct identification that did not yield high enough confidence; red denotes identification of the wrong serotype.

Neural Net 2 yielded less accurate results (Figure 4). Of the 208 spectra, only 63% of the total classification attempts were correct. Accuracy for all *Salmonella* 4,5,12,:i:-, Javiana, and Paratyphi B isolates was 0% while 100% accuracy occurred for all *Salmonella* Newport tested. The accuracy rate for *Salmonella* Kentucky isolates was 82%.

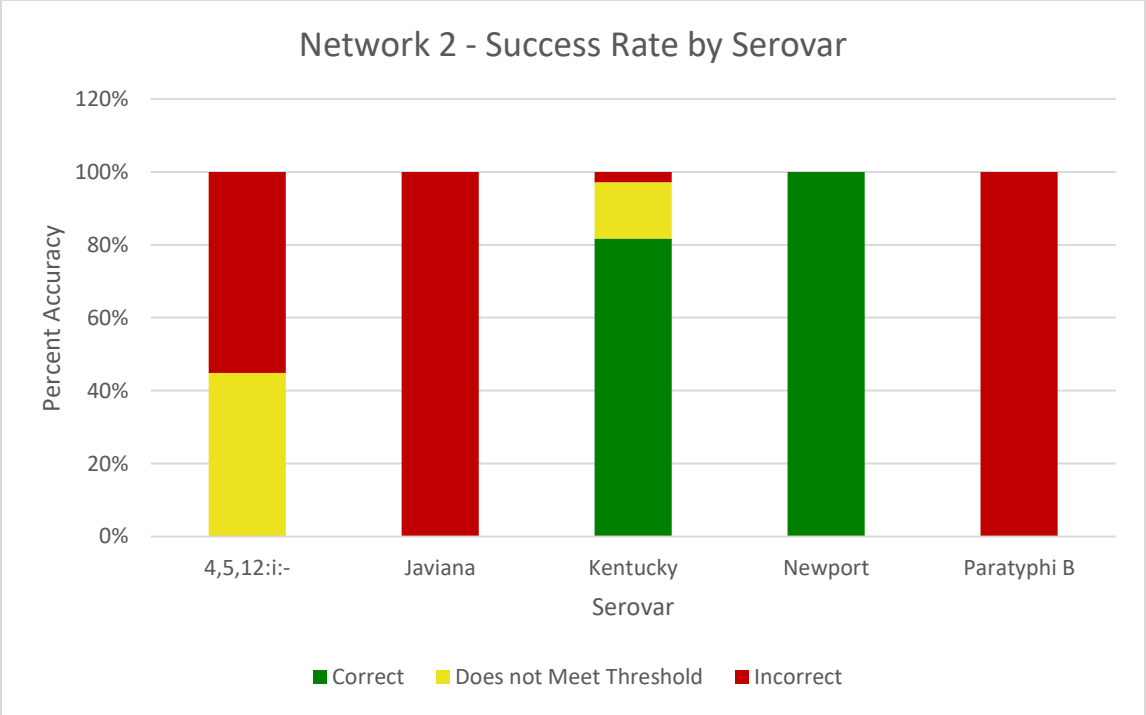


Figure 4: Success rate of Neural Net 2 organized by serovar. Green denotes correct identification; yellow denotes correct identification that did not yield high enough confidence; red denotes identification of the wrong serotype.

4 DISCUSSION

Within bacterial FTIR spectra are peaks/points within specific ranges that represent the presence of core biological building blocks: lipids (3000-2800 cm^{-1}), proteins/amides I and II (1700-1500 cm^{-1}), phospholipids/DNA/RNA (1500-1185 cm^{-1}), polysaccharides (1185-90 cm^{-1}), and the finger print region (900-600 cm^{-1}) (Maquelin et al., 2002 and Vaz et al., 2013). Between 2200 cm^{-1} and 2400 cm^{-1} is a large peak for all isolates prepared after March 24th, 2015. This anomaly could be a consequence of changes in the surrounding CO_2 levels in the laboratory, as the FTIR instrument can

easily detect changes in CO_2 concentration at even low quantities (Bruker Optics, 2010). CO_2 has a distinct peak between 2200 cm^{-1} and 2400 cm^{-1} that looks strikingly similar to the peaks found in this study. Despite this unexpected difference, the Neural Network analysis remained virtually unaffected due to the fact that the range between 1800 cm^{-1} and 2400 cm^{-1} were excluded from analysis.

While Neural Net 1 classifications yielded very accurate results, the second Neural Net classifications proved to be a much more difficult task for the ANN to correctly identify. While there is no clear answer yet, the glaring issue appears to be the sample sizes provided for these specific serotypes. The lack of isolates appears to have contributed to the lower number of unique points that could be processed into serotype features by the ANN in Neural Net 2. Three of the serovars deemed 'not separable' were from serovars that contained less than 30 spectra in total. It would be of great interest to acquire more of the five strains in order to build a more comprehensive library of the top 20 food-borne *Salmonella* serotypes.

A minor issue arises when considering the process in which each spectra is obtained. While the FTIR & ANN method together is considerably faster than slide agglutination and PFGE, it still relies on the standard methods in order to acquire confirmed isolates. FTIR & ANN analysis is certainly a rapid method for serotyping as long as the sample size is sufficient and is correctly analyzed to begin with. Once a strong reference library is built, the known serotypes can be easily detected within unknown samples.

Some serotypes used in this study did contain a sufficient number of isolates and were still calculated to be lacking in separation; particularly *Salmonella* Newport. While

Newport had 100% accuracy in Neural Net 2, when its spectra was introduced in the training of Neural Net 1, the failure rate dropped far below the average. Future research should investigate *Salmonella* Newport and its cell surface chemical composition in order to fully elucidate this phenomenon.

Salmonella Newport falls into three lineages when analyzed through a multilocus sequence typing scheme, each of which fall into multiple sequence types, some of which are pansusceptible to antibiotics. While not initially apparent, *Salmonella* Newport appears to be polyphyletic (Sangal et al., 2010), and very well may contain multiple lineages that have many variances in their cell surface composition, which may explain the disparity of accuracy when *S.* Newport is included with Net 1.

It is also apparent that, over time, Salmonellae among other bacteria are diversifying in response to the ever-changing environment. Genetic evidence shows that serotypes Agona, Kentucky, and Paratyphi B are also polyphyletic. *S.* Newport and Paratyphi B have also been shown to have multiple origins of lineage. It is suggested that the genes responsible for O groups and flagellar antigen are not evolving in a linear fashion (Timme et al., 2013). While this is not an immediate concern, the daunting task of tracking and classifying *Salmonella* serotypes that are evolving in a non-linear fashion looms in the future.

While slide agglutination is an excellent method of serotyping, FTIR drastically improves on the time and resources needed to achieve the same goals. Though, without slide agglutination or PFGE, Fourier-Transform Infrared Spectroscopy as a method for serotyping is not reliable. With a master neural network shared globally, laboratories around the world could utilize and continue to update the amount and quality of spectra

with a constantly evolving and growing artificial neural network that can ideally identify thousands of *Salmonella* strains accurately. This study was done with the resources of a single microbiology laboratory; the ubiquitous presence of coherence through technology will inevitably allow collaboration and reproducibility to reach the upper limits of efficiency.

5 REFERENCES CITED

Abrahams, G. L. & Hensel, M., (2006). Manipulating cellular transport and immune responses: dynamic interactions between intracellular *Salmonella* enterica and its host cells. *Cell Microbiology,* 8(5):728-737.

Ahmer, B. & Gunn, J., (2011). Interaction of *Salmonella* spp. with the intestinal microbiota. *Frontiers in Microbiology,* 2(101).

Barco, L., Longo, A., Lettini, A. A., Cortini, E., Saccardin, C., Minorello, C., Olsen, J. E. & Ricci, A., (2014). Molecular Characterization of ‘‘Inconsistent’’ Variants of *Salmonella* Typhimurium Isolated in Italy. *Foodborne Pathogens and Disease* 11(6):497-499.

Barman, M., Unold, D., Shifley, K., Amir, E., Hung, K., Bos, N. & Salzman, N., (2008). Enteric Salmonellosis Disrupts the Microbial Ecology of the Murine Gastrointestinal Tract. *Infection and Immunity,* 76(3):907-915.

Baron, S. (1996). Medical Microbiology 4th Edition Ch. 21, Salmonella.

Basheer, I., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods, 43*(1), 3-31.

Bertrand, S., Dierick, K., Heylen, K., De Baere, T., Pochet, B., Robesyn, E., & Lokietek, S. (2010). Lessons Learned from the Management of a National Outbreak of *Salmonella* Ohio Linked to Pork Meat Processing and Distribution. *Journal of Food Protection*.

Bruker Optics, (2010) Application Note AN 107, FT-IR Spectroscopy in Ultrahigh vacuum: Surface Science Approach for Understanding Reactions on catalytic Oxide Powders. BOPT-4000518 – 01.

CDC, (1968-2011). An Atlas of *Salmonella* in the United States, Available at: http://www.cdc.gov/Salmonella/pdf/Salmonella-atlas-508c.pdf.

CDC, (2012). National Enteric Disease Surveillance: *Salmonella* Annual Report, Available at: http://www.cdc.gov/ncezid/dfwed/pdfs/*Salmonella*-annual-report-2012-508c.pdf.

CDC, (2013). Standard Operating Procedure for PulseNet PFGE of *Escherichia coli* O157:H7, *Escherichia coli* non-O157 (STEC), *Salmonella* serotypes, *Shigella* sonnei and *Shigella* flexneri Available at: http://www.cdc.gov/pulsenet/pdf/ecoli-shigella-salmonella-pfge-protocol-508c.pdf.

Deiwick, J., Salcedo S. P., Boucrot, E., Gilliland, S. M., Henry, T., Petermann, N., . . . Meresse, S., (2006). The translocated *Salmonella* effector proteins SseF and SseG interact and are required to establish an intracellular replication niche. *Infection and Immunity* 74(12):6965-72.

Difco Laboratories, (1998). The Difco Manual (11th ed., pp. 664-675). Sparks, MD: Difco Laboratories, Division of Becton Dickinson and Co.

Gal-Mor, O., Boyle, E., & Grassl G., (2014). Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ. *Frontiers in Microbiology, 2014; 5: 391*.

Garai, P., Gnanadhas, D. P., & Chakravortty, D., (2012). *Salmonella* enterica serovars Typhimurium and Typhi as model organisms: Revealing paradigm of host-pathogen interactions. *Virulence* 3(4):377-388.

Gonzalez-Escobedo, G., Marshall, J., & Gunn, J., (2011). Chronic and acute infection of the gall bladder by *Salmonella* Typhi: understanding the carrier state. *Nature Reviews Microbiology*, 9(1), 9-14

Grimont, P., & Weill, F. (2007). Antigenic Formulae of the Salmonella Serovars. *WHO Collaborating Centre for Reference and Research on Salmonella*.

Groisman, E., Ochman, H., (1996). Pathogenicity Islands: Bacterial Evolution in Quantum leaps. *Cell*, 87(5):791-794.

Guibourdenche, M., Roggentin, P., Mikoleit, M., Fields, P. I., Bockemühl, J., Grimont, P. A., & Weill, F. (2010). Supplement 2003–2007 (No. 47) to the White-Kauffmann-Le Minor scheme. *Research in Microbiology*, *161*(1), 26-29.

Hallstrom, K. & McCormick, B., (2011). *Salmonella* Interaction with and Passage through the Intestinal Mucosa: Through the Lens of the Organism. *Frontiers in Microbiology*, 2(88).

Ibarra, J. A. & Steele-Mortimer, O., (2009). *Salmonella* – the ultimate insider. *Salmonella* virulence factors that modulate intracellular survival. *Cellular Microbiology*, 11(11): 1579-1586.

Ibarra, J. A., Knodler, L. A., Sturdevant, D. E., Virtaneva, K., Carmody, A. B., Fischer, E. R., . . . Steele-Mortimer, O., (2010). Induction of *Salmonella* pathogenicity island 1 under different growth conditions can affect *Salmonella*-host cell interactions *in vitro*. *Microbiology.* 156(Pt 4): 1120-1133.

Kérouanton, A., Marault, M., Lailler, R., Weill, F., Feurer, C., Espié, E., & Brisabois, A. (2007). Pulsed-Field Gel Electrophoresis Subtyping Database for Foodborne *Salmonella enterica* Serotype Discrimination. *Foodborne Pathogens and Disease, 4*(3), 293-303.

Kim, S., Reuhs, B., & Mauer, L. (2005). Use of Fourier Transform Infrared Spectra of Crude Bacterial Lipopolysaccharides and Chemometrics for Differentiation of *Salmonella enterica* serotypes. *J Appl Microbiol Journal of Applied Microbiology*, *99*(2): 411-417.

Koenig, J., (1981) Fourier Transform Infrared Spectroscopy of Chemical Systems. *Accounts of Chemical Research* 14: 171-178.

Lamprell, H., Mazerolles, G., Kodjo, A., Chamba, J. F., Noel, Y., & Beuvier, E., (2006) Discrimination of *Staphylococcus aureus* strains from different species *Staphylococcus* using Fourier Transform infrared (FTIR) spectroscopy. *International Journal of Food Microbiology* 108(1): 125-129.

Lupp, C., Robertson, M., Wickham, M., Sekirov, I., Champion, O., Gaynor, E., & Finlay B., (2007). Host-Mediated Inflammation Disrupts the Intestinal Microbiota and

Promotes the Overgrowth of Enterobacteriaceae. *Cell Host & Microbe*, 2(2):119-129.

Maquelin, K., Kirschner, C., Choo-Smith, L., Braak, N. V., Endtz, H., Naumann, D., & Puppels, G. (2002). Identification of medically relevant microorganisms by vibrational spectroscopy. *Journal of Microbiological Methods, 51*(3), 255-271.

Romanolo, K. F., Gorski, L., Wang, S., & Lauzon, C. R., (2015). Rapid Identification and Classification of *Listeria* spp. and Serotype Assignment of *Listeria monocytogenes* Using Fourier Transform-Infrared Spectroscopy and Artificial Neural Network Analysis *PLoS ONE* 10(11): e0143425.

Sangal, V., Harbottle, H., Mazzoni, C. J., Helmuth, R., Guerra, B., Didelot, X., & Achtman, M. (2010). Evolution and Population Structure of *Salmonella enterica* Serovar Newport. *Journal of Bacteriology, 192*(24), 6465-6476.

Scallan, E. (2011). Foodborne Illness Acquired in the United States (Reply). *Emerg. Infect. Dis. Emerging Infectious Diseases, 17*(7), 1339-1340.

Schrader, K. N., Fernandez-Castro, A., Cheung, W. K., Crandall, C. M., & Abbott, S. L. (2008). Evaluation of Commercial Antisera for *Salmonella* Serotyping. *Journal of Clinical Microbiology, 46*(2), 685-688.

Schwartz, C. D. & Cantor, R. C. (1984). Separation of Yeast Chromosome-Sized DNAs by Pulse Field Gradient Gel Electrophoresis. *Cell*, 37, 67-75.

Sirinavin, S., Jayanetra, P., & Thakkinstian, A. (1999). Clinical and Prognostic Categorization of Extraintestinal Nontyphoidal *Salmonella* Infections in Infants and Children. *Clinical Infectious Diseases, 29*(5), 1151-1156.

Steele-Mortimer, O., Meresse, S., Gorvel, J. P., Toh & B.H.,Finlay, B. B., (1999). Biogenesis of *Salmonella* typhimurium-containing vacuoles in epithelial cells involves interactions with the early endocytic pathway. *Cellular Microbiology* 1: 33–49.

Threlfall, E., Hampton, M., Ward, L., Richardson, I., Lanser, S., & Greener, T. (1999). Pulsed field gel electrophoresis identifies an outbreak of *Salmonella enterica* serotype Montevideo infection associated with a supermarket hot food outlet. *COMMUNICABLE DISEASE AND PUBLIC HEALTH, 2*(3), 207-209.

Timme, R. E., Pettengill, J. B., Allard, M. W., Strain, E., Barrangou, R., Wehnes, C., . . . Brown, E. W. (2013). Phylogenetic Diversity of the Enteric Pathogen *Salmonella enterica* subsp. *enterica* Inferred from Genome-Wide Reference-Free SNP Characters. *Genome Biology and Evolution, 5*(11), 2109-2123.

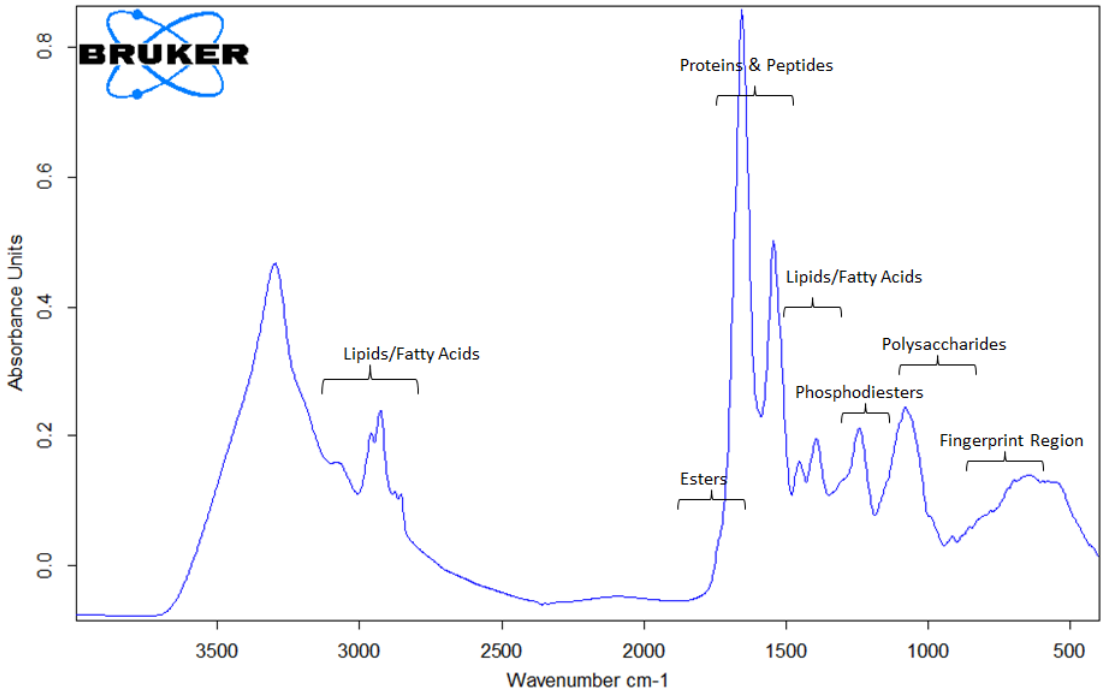
Vaz, M., Meirinhos-Soares, L., Sousa, C., Ramirez, M., Melo-Cristino, J., & Lopes, J. (2013). Serotype discrimination of encapsulated Streptococcus pneumoniae strains by Fourier-transform infrared spectroscopy and chemometrics. *Journal of Microbiological Methods, 93*(2), 102-107.

Wattiau, P., Boland, C., & Bertrand, S. (2011). Methodologies for *Salmonella enterica* subsp. *enterica* Subtyping: Gold Standards and Alternatives. *Applied and Environmental Microbiology, 77*(22), 7877-7885.

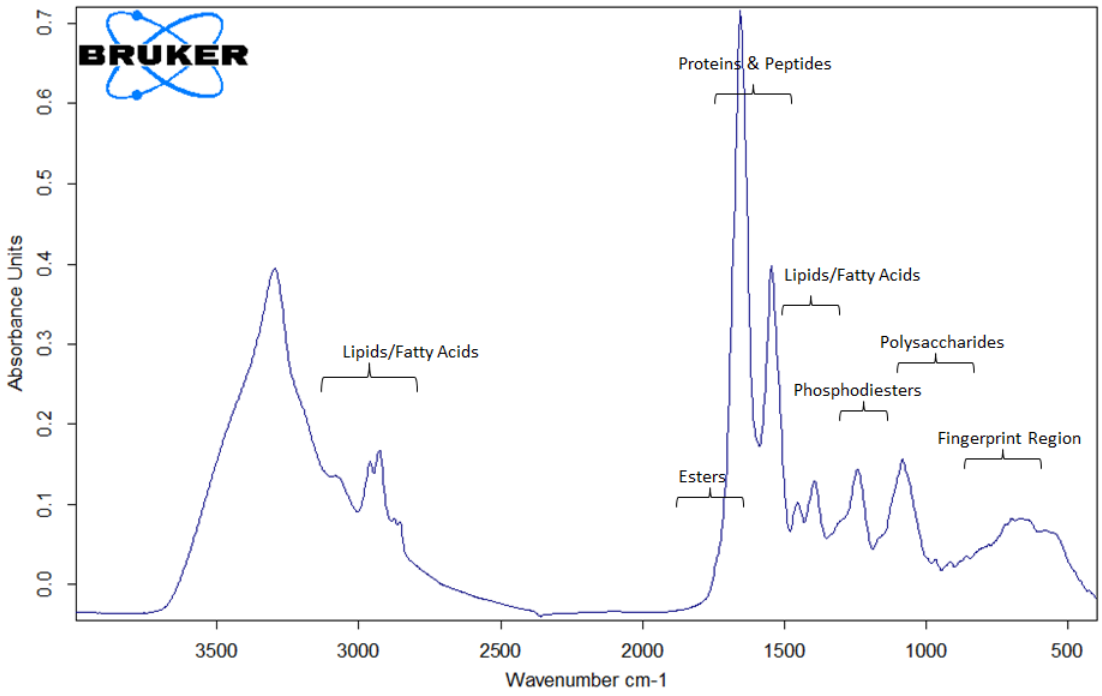
6 APPENDIX

Example spectra for each serotype, organized by date acquired. Peaks between 2200 cm^{-1} and 2400 cm^{-1} can be observed starting with *Salmonella* Javiana, April 3rd, 2015.

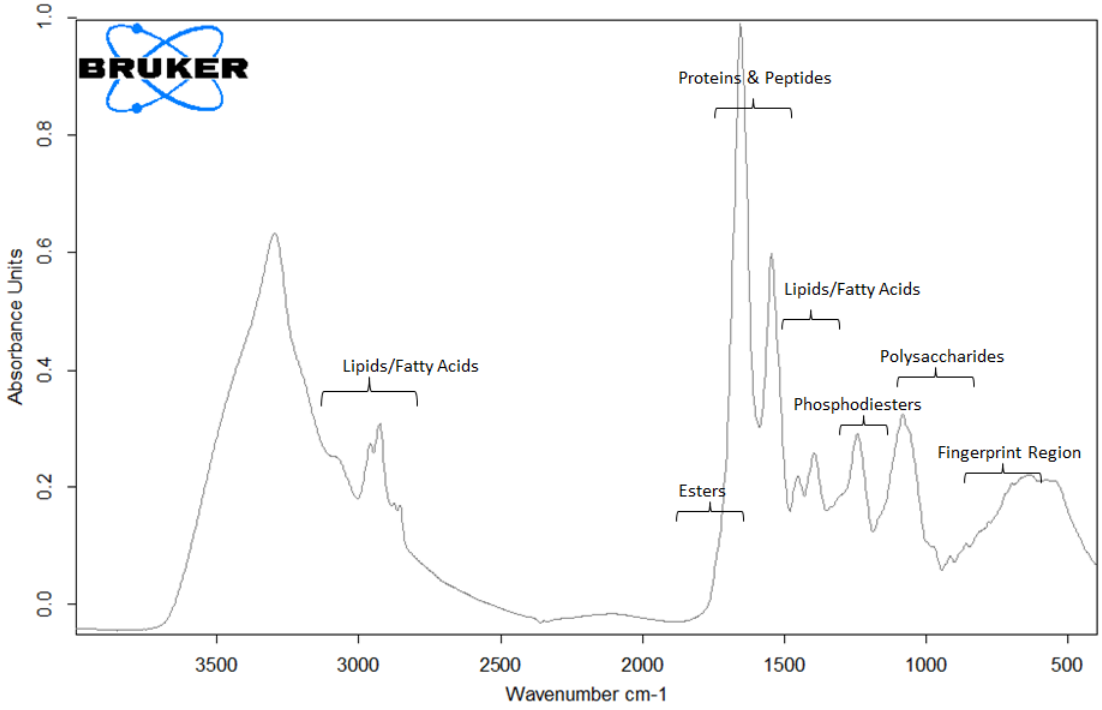
March 4th, 2015 - *Salmonella* Enteritidis



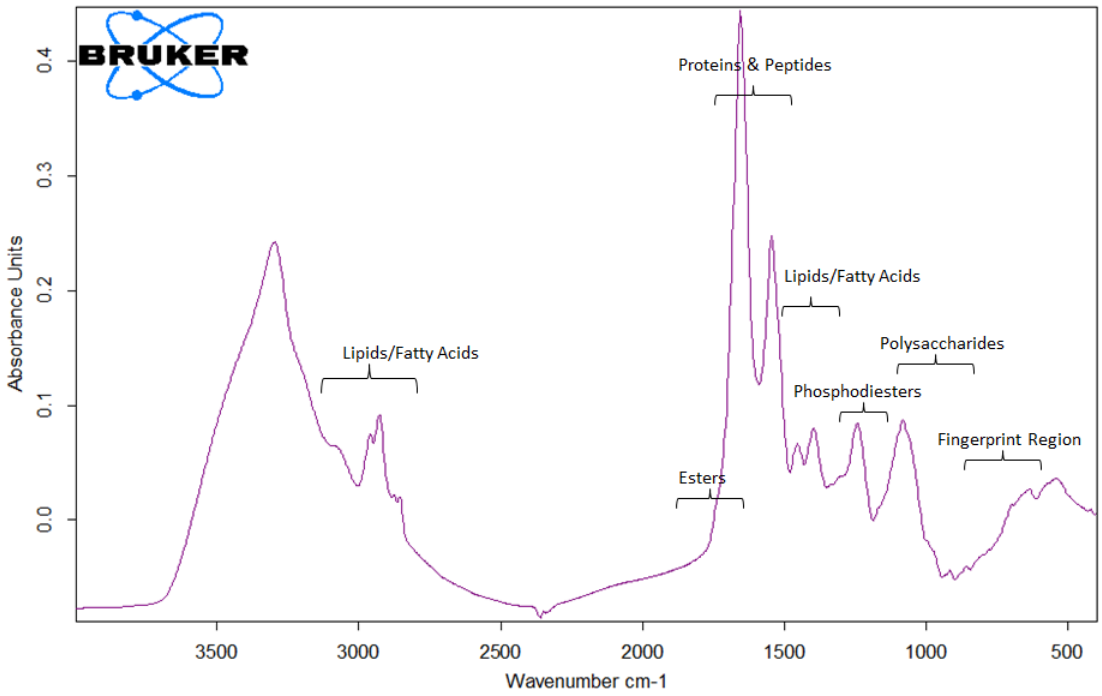
March 4th, 2015 – *Salmonella* Newport



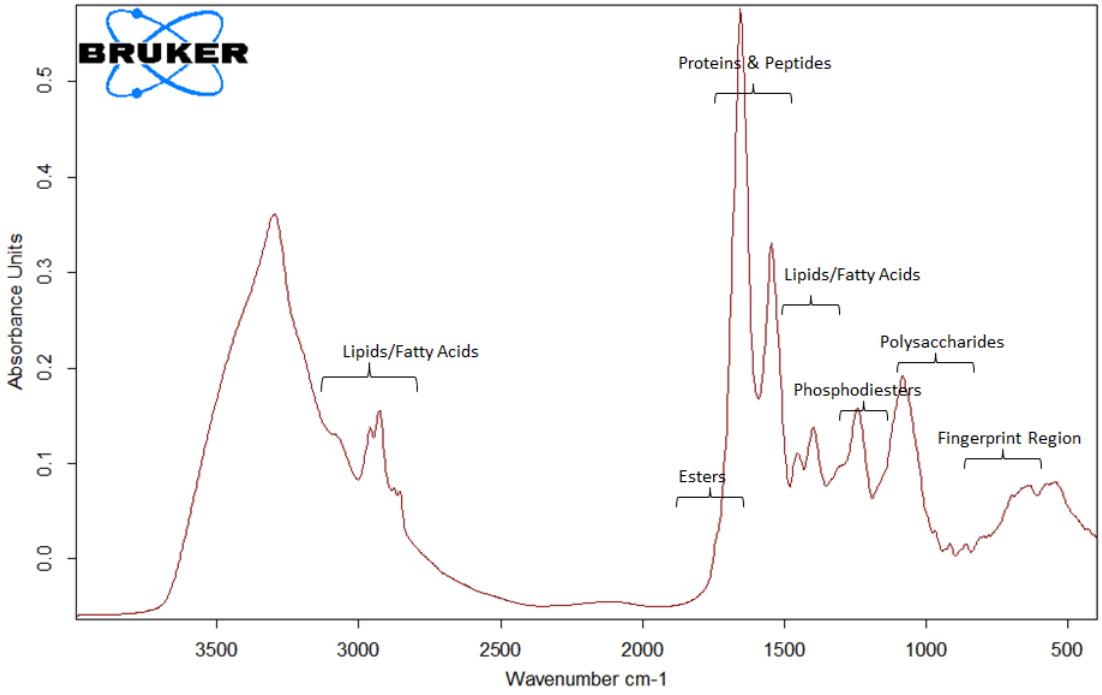
March 4th, 2015 – *Salmonella* Typhimurium



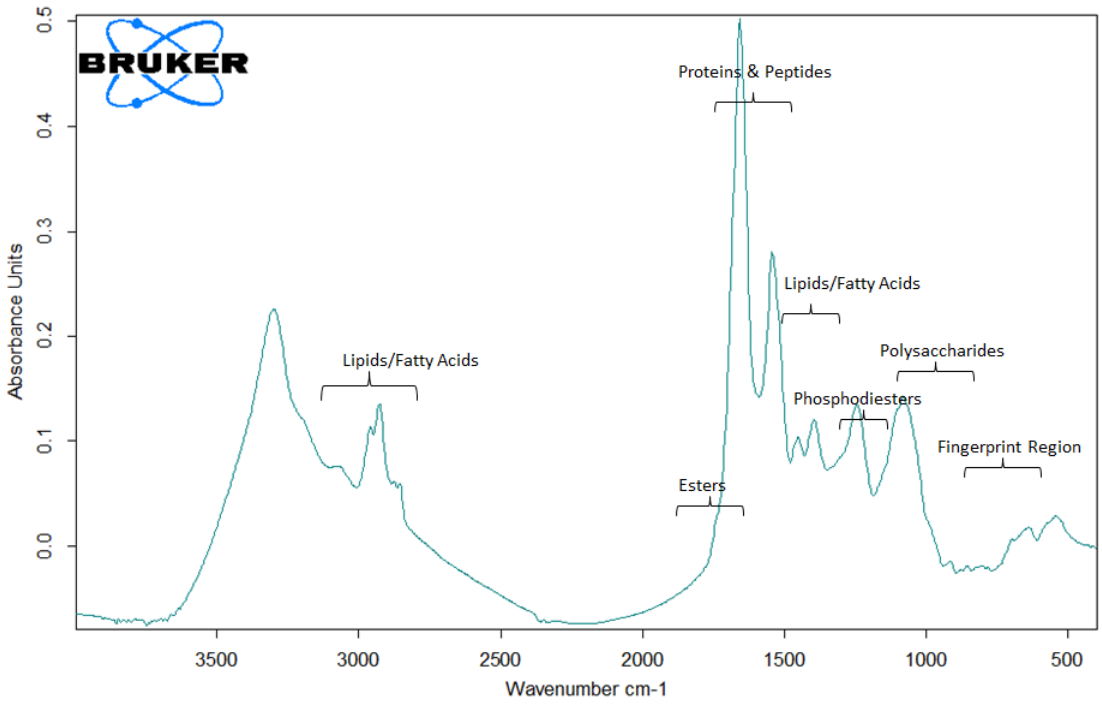
March 13th, 2015 – *Salmonella* Heidelberg



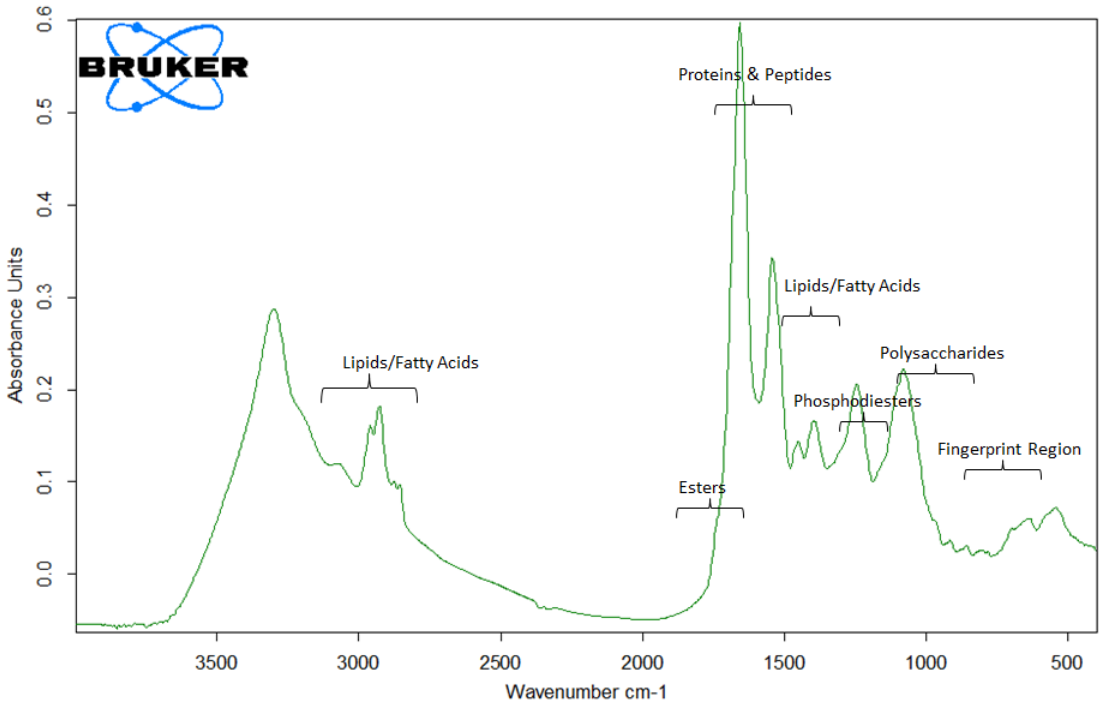
March 13th, 2015 – *Salmonella* Montevideo



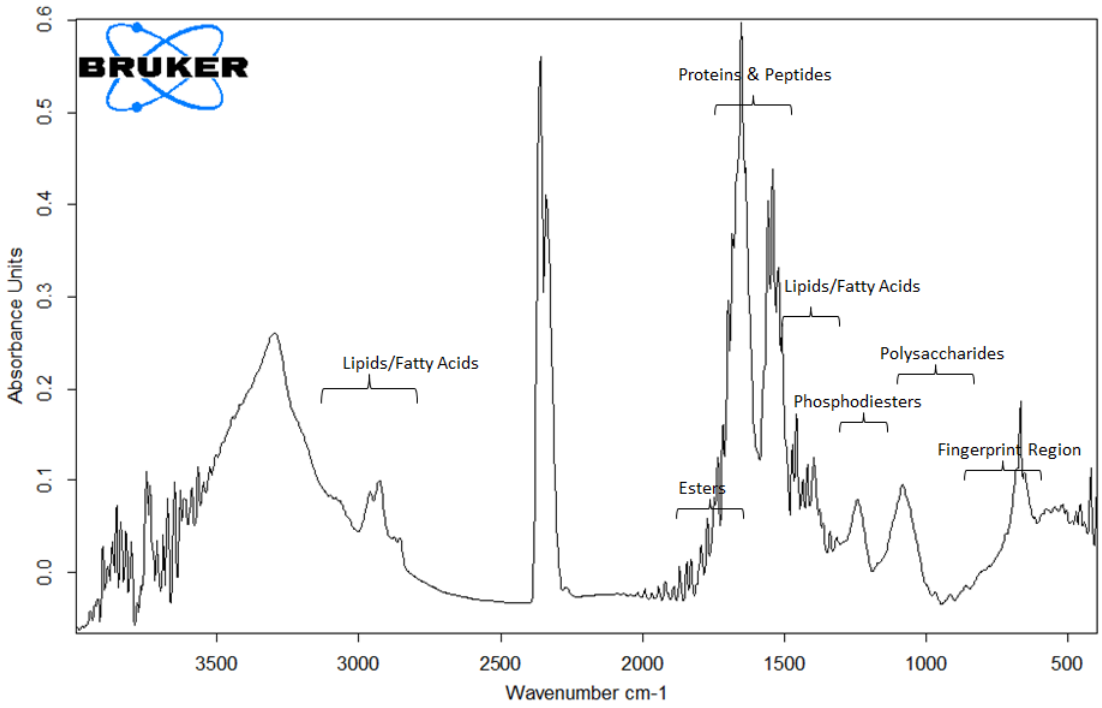
March 13th, 2015 – *Salmonella* Javiana



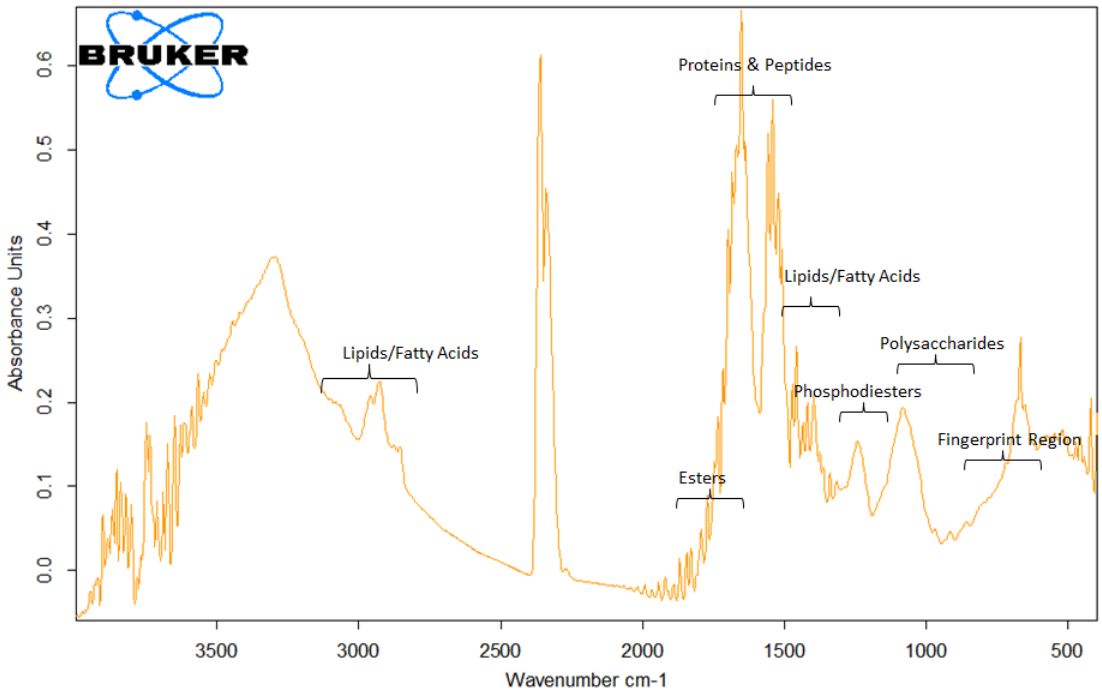
March 24th, 2015 – *Salmonella* Muenchen



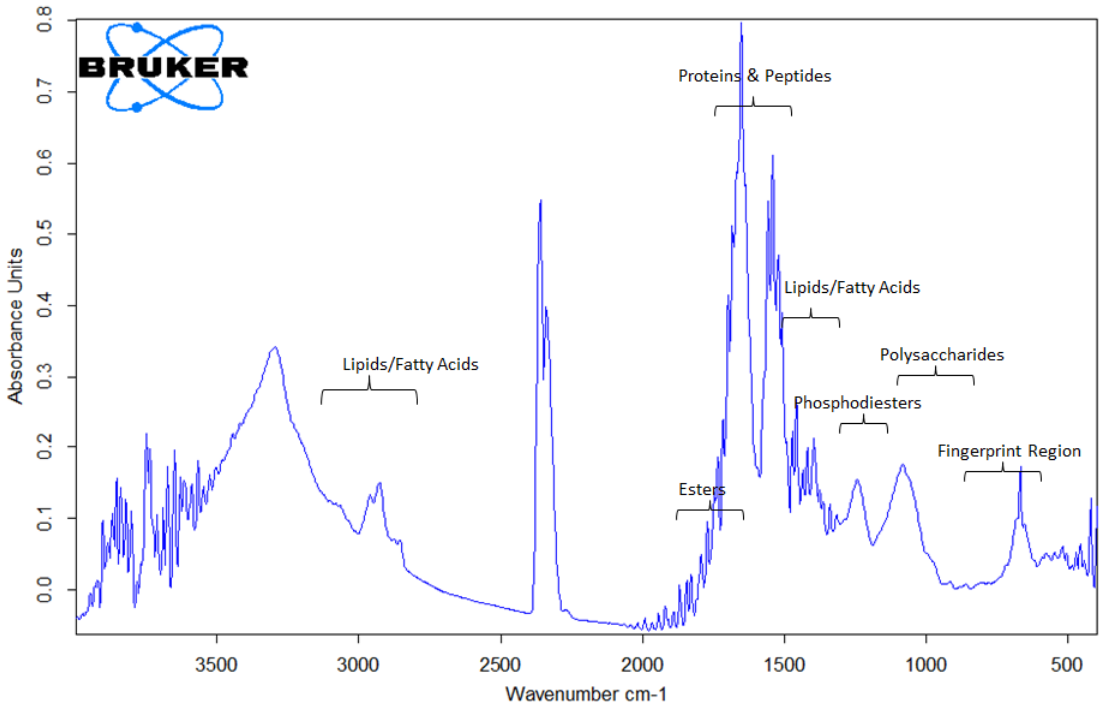
April 3rd, 2015 – *Salmonella* Infantis



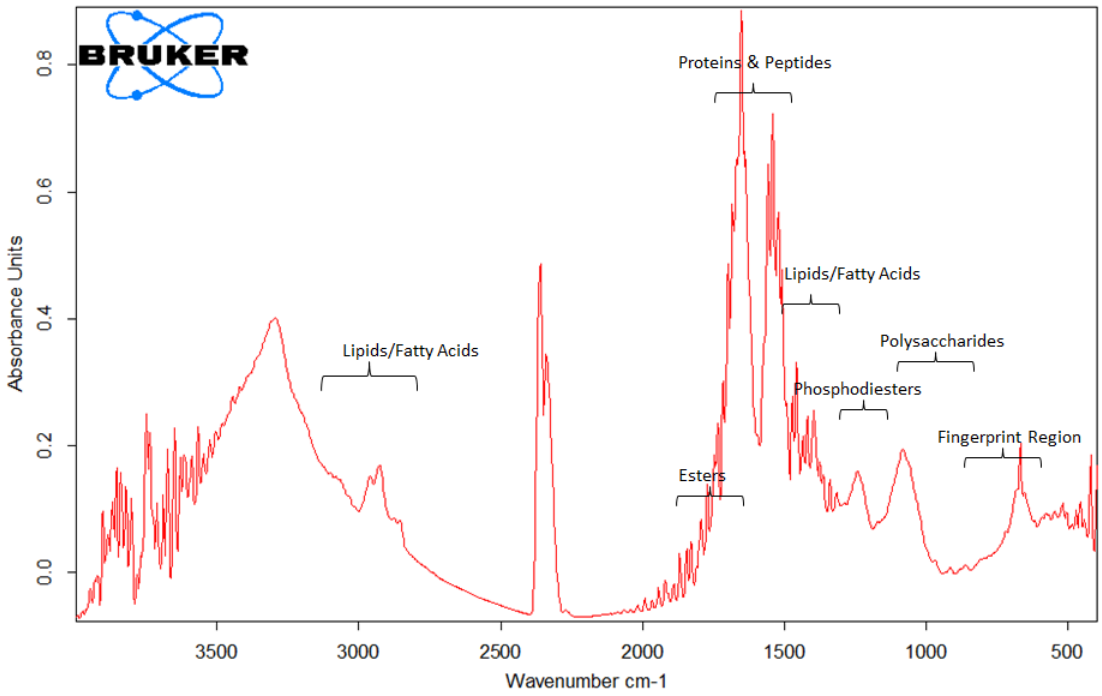
April 10th, 2015 – *Salmonella* Oranienberg



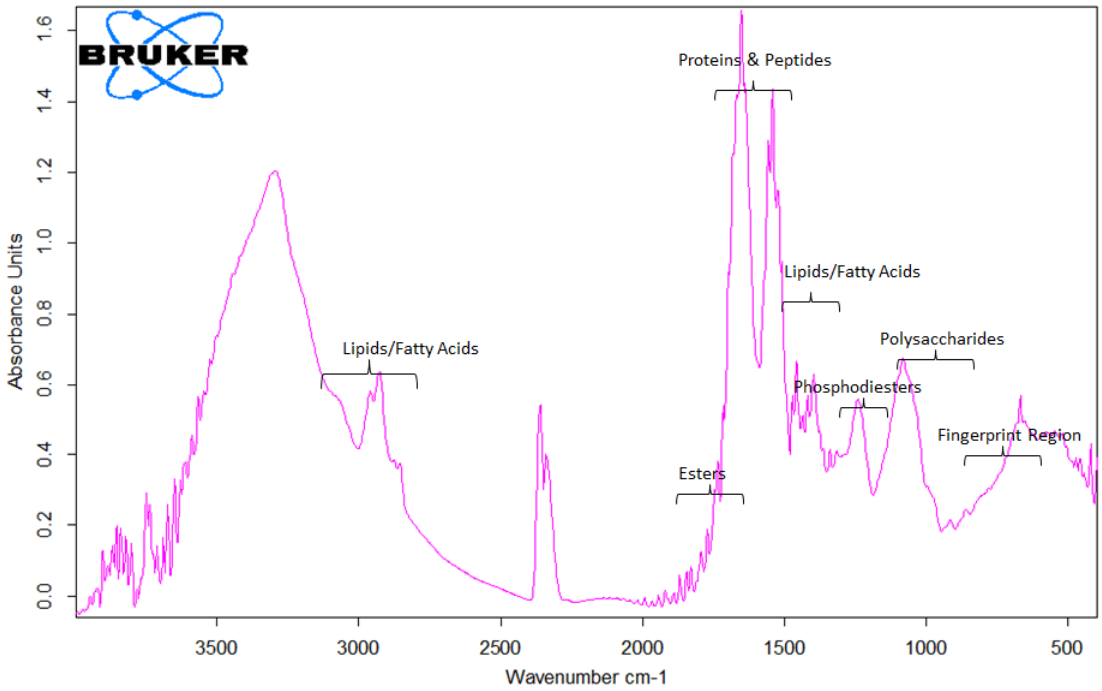
April 16th, 2015 – *Salmonella* Saintpaul



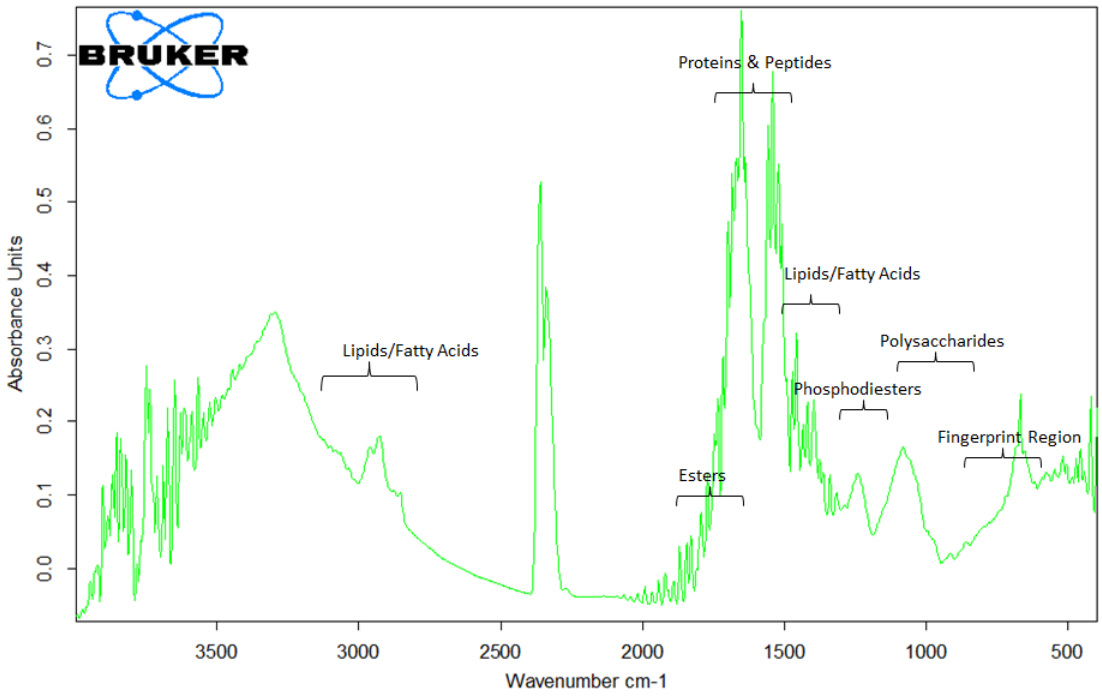
April 24th, 2015 – *Salmonella* Thompson



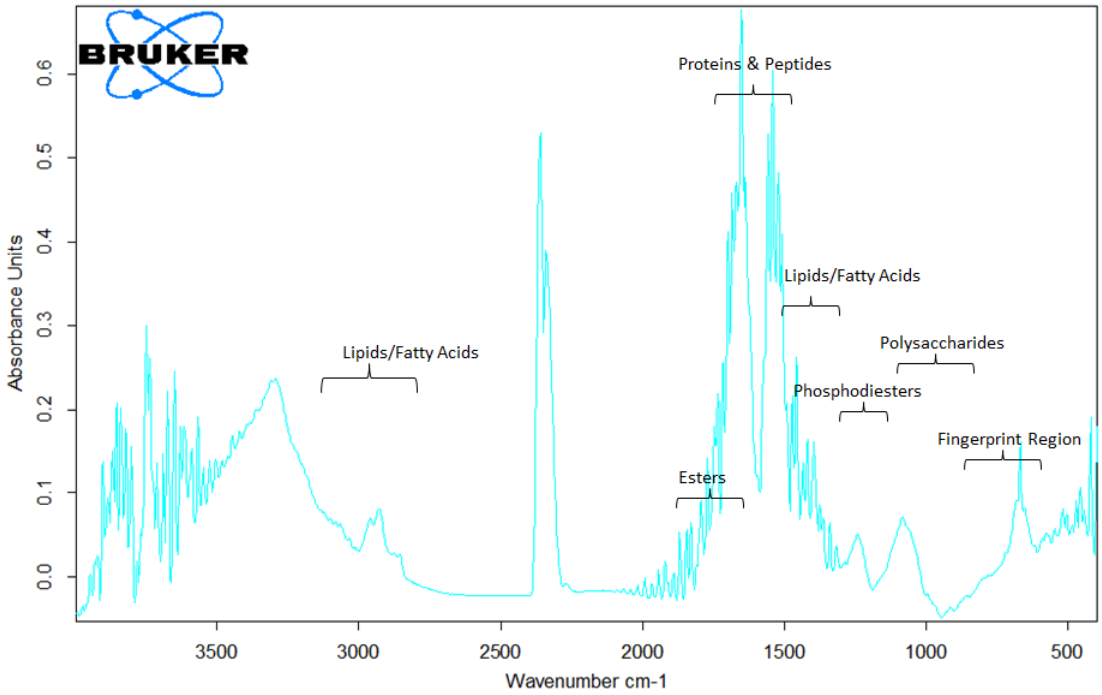
May 15th, 2015 – *Salmonella* 4,5,12,:i:-



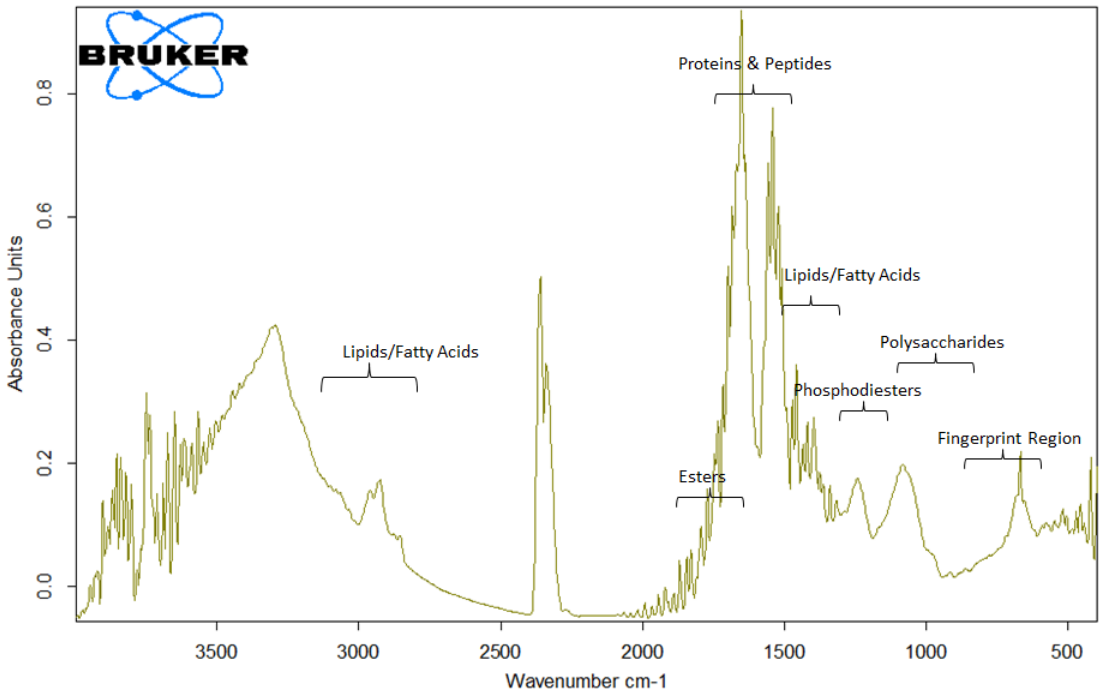
May 15th, 2015 – *Salmonella* Agona



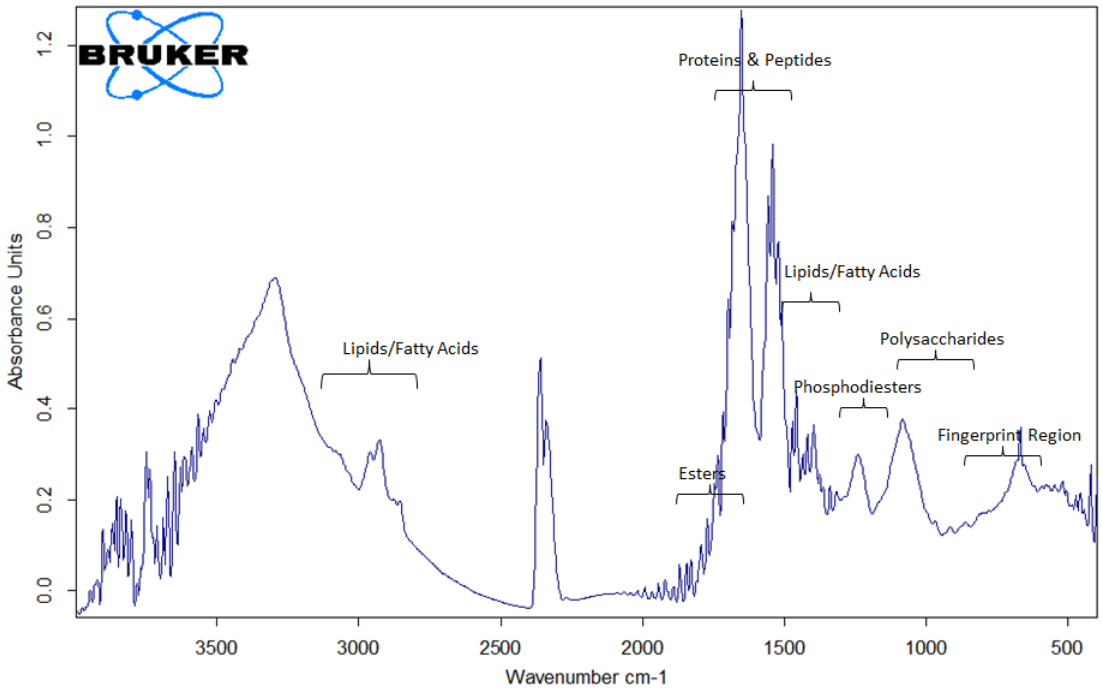
June 1st, 2015 – *Salmonella* Anatum



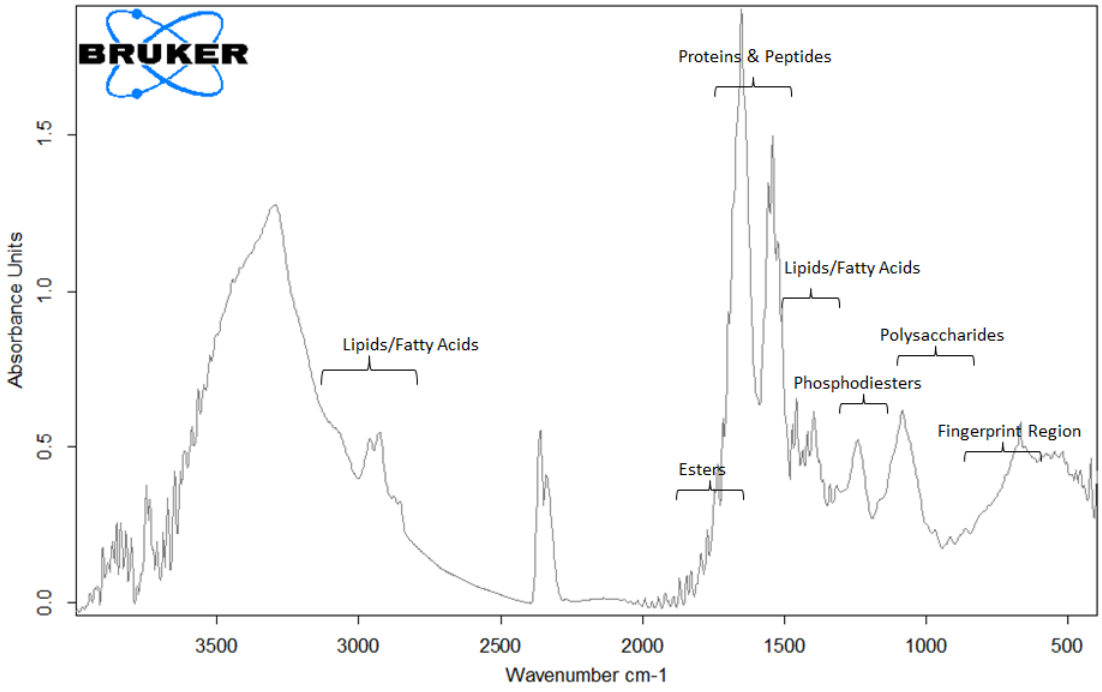
June 5th, 2015 – *Salmonella* 4,5,12,:i:-



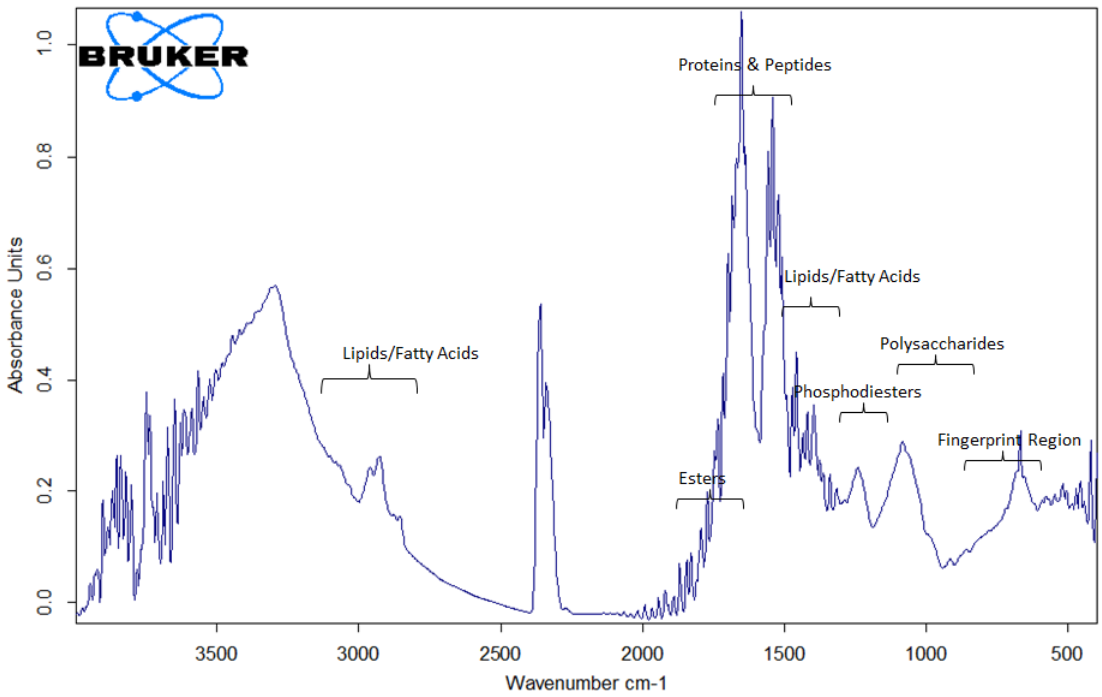
June 5th, 2015 – *Salmonella* Braenderup



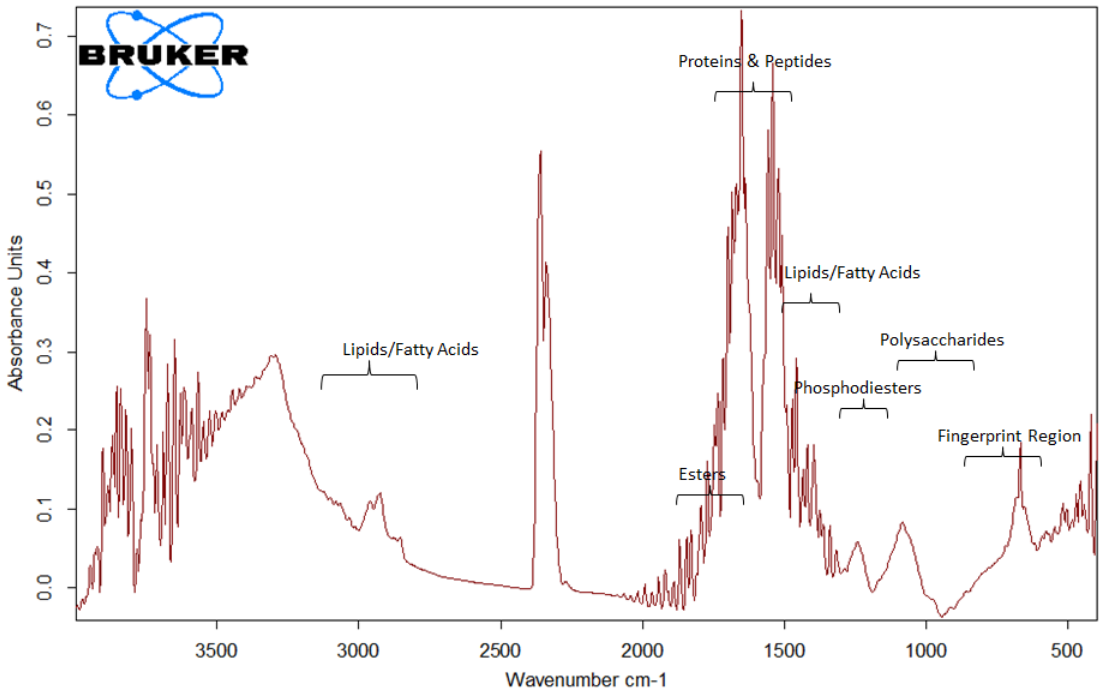
July 10th, 2015 – *Salmonella* Berta



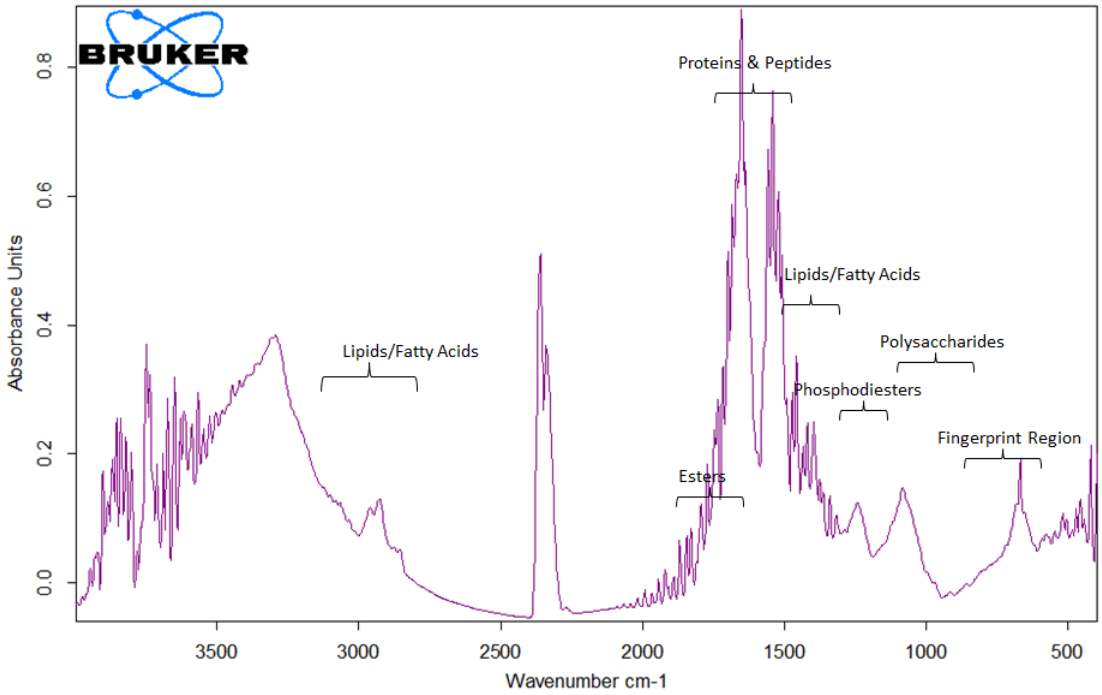
July 10th, 2015 – *Salmonella* Javiana



July 10th, 2015 – *Salmonella* Paratyphi B



July 17th, 2015 – *Salmonella* Kentucky


BRUKER
Proteins & Peptides
Lipids/Fatty Acids
Polysaccharides
Phosphodiesters
Fingerprint Region
Esters
Lipids/Fatty Acids
Absorbance Units
0.0
0.2
0.4
0.6
0.8
3500
3000
2500
2000
1500
1000
500
Wavenumber cm-1