UNIVERSITY OF CALIFORNIA, SAN DIEGO

Advanced High Resolution Melt Technology for Bacteria Serotyping and Antibiotic Resistance Detection

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Bioengineering

by

Yang Zhang

Committee in charge:

Professor Stephanie Fraley, Chair
Professor Shu Chien
Professor Adam Engler

2016

The thesis of Yang Zhang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

DEDICATION

To thank my parents and friends without whom I will never imagine what I can be today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Dr. Stephanie Fraley, Sinead Hawker, Daniel Ortiz, Aarthi Saminathan, Hannah Mack, Ninad Kulkarni, Behnam Hedayatnia, Yijia Pan, Ye Gao, Shengye Wang.

ABSTRACT OF THE THESIS

Advanced High Resolution Melt Technology for Bacteria Serotyping and Antibiotic Resistance Detection

by

Yang Zhang

Master of Science in Bioengineering

University of California, San Diego, 2016

Professor Stephanie Fraley, Chair

Bacteria identification plays an important role in confirming diagnosis of bacterial disease and tracing the source of an outbreak for public health. However, the standard method for bacteria identification which relies on the colony feature of the bacteria cultured from the patient's specimen is time consuming and lack of accuracy and sensitivity. New technology such as qPCR based sequencing and microarray showed

their benefit in this area but are still suffering problems regarding to cost, complexity and accuracy. Therefore, this thesis tries to advance the high resolution melt (HRM) technology, a molecular technique to detect difference in double-stranded DNA, to realize bacteria serotyping. Two methods: intercalating dyes and multiplex PCR were investigated for this purpose. To our excitement, multiplex PCR of V1 to V6 region and V6 region of the gene 16S ribosome RNA of Salmonella followed by HRM can distinguish 5 serotypes of Salmonella under 3 hours, including Newport, Typhimurium, Heidelberg, Enteritidis and Choleraesuis. We further applied this technology to achieve antibiotic resistance detection and bacteria speciation in one test and successfully detected the kanamycin resistance from E. coli. This multiplex PCR combined with HRM will be further combined to digital PCR to build a fast, easy to use, highly specific, highly sensitive and absolute quantitative tool for bacteria diagnosis.

CHAPTER 1: Introduction

1.1 Bacteria identification

1.1.1 Needs for serotype identification

Bacteria identification and serotypes identification are needed for two main reasons. First of all, rapid identification of pathogenic bacteria is crucial for improving patient care. Take Salmonella as an instance. Different serotypes of Salmonella might require different treatment. Serotype Choleraesuis is more aggressive and once infected might cause other infection along with the food poisoning. Therefore, patient with Choleraesuis infection needs not only the rehydration treatment to stop the symptom but also antibiotic to prevent the further spreading of the bacteria. Therefore, to identify the serotype can provide the information assisting the doctor to give a more relevant medication.

Second, species and serotypes identification is a must in public health to trace back the source and to predict or prevent the outbreak of infectious disease. Since 1960, public health scientists in the US have used serotyping to help find Salmonella outbreaks and track them back to their source (1). In public health, if a patient is suspect of Salmonella infection, laboratory scientists will first identify this infection by culturing the patient's sample to "confirm the culture". Then the diagnostic laboratories report this result to the treating clinician and submit Salmonella isolates to state and territorial public health

laboratories for serotyping and DNA fingerprinting. After that, the public health laboratories will report the results to Centers for Disease Control (CDC) to digitally mapping the source of this case with other cases to monitor the disease occurrence and potential for any outbreak.

1.1.2 Bacteria identification in clinical laboratory

In clinics, the identification is based on taxonomic principles applied to the clinical microbiological situation (2). Bacteria culture and biochemical staining remain the gold standard despite its long procedural time (3-5 days) and limitation in accuracy. Specifically, the bacteria from patient stool or blood sample is cultured overnight on agar plates, and then the clinical laboratory scientist will classify the organism based on morphological and metabolic characteristics, such us the bacteria colony color, shape and etc. A series of biochemical test can also be conducted to confirm the genus and species. As this diagnostic test largely depends on discriminating information given from the bacteria and the empirical experience of the scientist, it is limited in identifying certain species not to mention serotype identification.

1.2 New technology for bacteria detection

1.2.1 Arising of new technology

In the past decades, various technologies were explored to satisfy this unmet need in clinical diagnosis, including amplification, hybridization and mass spectrometry, even though no clinical outcome trials have been conducted to prove in which the use of these novel methods would guide antimicrobial therapy (3).

PCR and microarray based molecular tools are most investigated for this purpose. Real time PCR combined with sequencing can detect bacteria from small amounts of cultured bacteria cells. Because primers can be designed to target bacteria at any level of specificity: genus, species or serotypes, this technology improves the sensitivity and decreases the time required for bacteria identification.

Microarrays has the potential of simultaneous bacterial identification and speciation. The designing of species-specific oligonucleotide probes to target different bacteria enables the rapid detection of multiple bacteria on one slide. This is extremely important for clinical samples. However, both real time PCR and microarray lack the sensitivity to detect individual sequence at low concentration level. Meanwhile, the highly specific primer or probe annealing often results in inaccuracy, especially for long sequence.

Meanwhile, mass spectrometry (MS) and nanotechnology were also employed to circumvent the problem of procedural time and detection limit. Immunoaffinity mass spectrometry applies the MS technology to detect the bacteria extracted by antibodies-modified magnetic beads can reach a detection resolution of 500 cells per ml in blood serum (4). Nanoparticles and oligonucleotide probes designed to capture the single-strand bacteria DNA following asymmetric reverse transcription-PCR and detected by a miniaturized uNMR device can simultaneously diagnosing a panel of 13 bacterial species in clinical specimens within 2 h (5).

Recently, more rapid molecular diagnosis assays for sepsis diagnosis have been introduced and evaluated (6); including LightCycler SeptiFast Test (7), peptide nucleic acid fluorescence in situ hybridization (PNA-FISH) (8), and matrix-assisted laser desorption-ionization time of flight mass spectrometry (MALDI-TOF MS) (9), and a DNA-based microarray platform [Prove-it sepsis assay (10) and the Verigene Gram-Positive Blood Culture (BC-GP) assay (11).

Among all of these technologies, the Verigene Gram-Negative Blood Culture (BC-GN) Nucleic Acid Test (Nanosphere Inc., Northbrook, IL) is a technology recently approved by the U.S. Food and Drug Administration. It is a sample-to-result automated microarray-based, multiplexed assay for species identification of GNB and detection of

their drug resistance genes in positive blood culture bottles, and is designed to directly detect species of GNB from positive blood culture bottles with 5 minutes of hands-on and 2 hours of run time per sample (12).

Besides all the above mentioned technologies, there is another PCR based technology has great potential to provide high sensitivity, high accuracy, short time and low cost test for bacteria identification and is intensively explored by our lab, the high resolution melt (HRM).

1.2.2 High resolution melt

High resolution melt (HRM) is a PCR based technology to characterize the sequence of DNA according to its own physical property. This process usually follows real-time PCR (qPCR) in which an intercalating dye has already been added as the fluorescent indicator. The intercalating dye specifically binds to double-strand DNA and form the stable fluorescent and thus can monitor the relative quantity of the product during DNA amplification in qPCR. High resolution melt then uses the intercalating dye to monitor the change of DNA from double strand to single strand. Specifically, the reaction product will be heated to 95 degree which allows the double strands to open and the dye to bind. Then, the sample will be cool down to 45 degree (can vary depends on needs) and the

double strands will rewind with the dye firmly embedded between the two strands and give out fluorescent signal. After this a slow heating up step follows. The gradually increased temperature will cause the double strands to open and intercalating dye leave the DNA but in a slower speed. Until the temperature reaches more than 90 degree, the double-stranded DNA will completely open and completely lose fluorescence. Therefore, in this process we will get a fluorescent dropping curve indicating the double-stranded DNA is changing to single strand DNA shown in Figure 1.

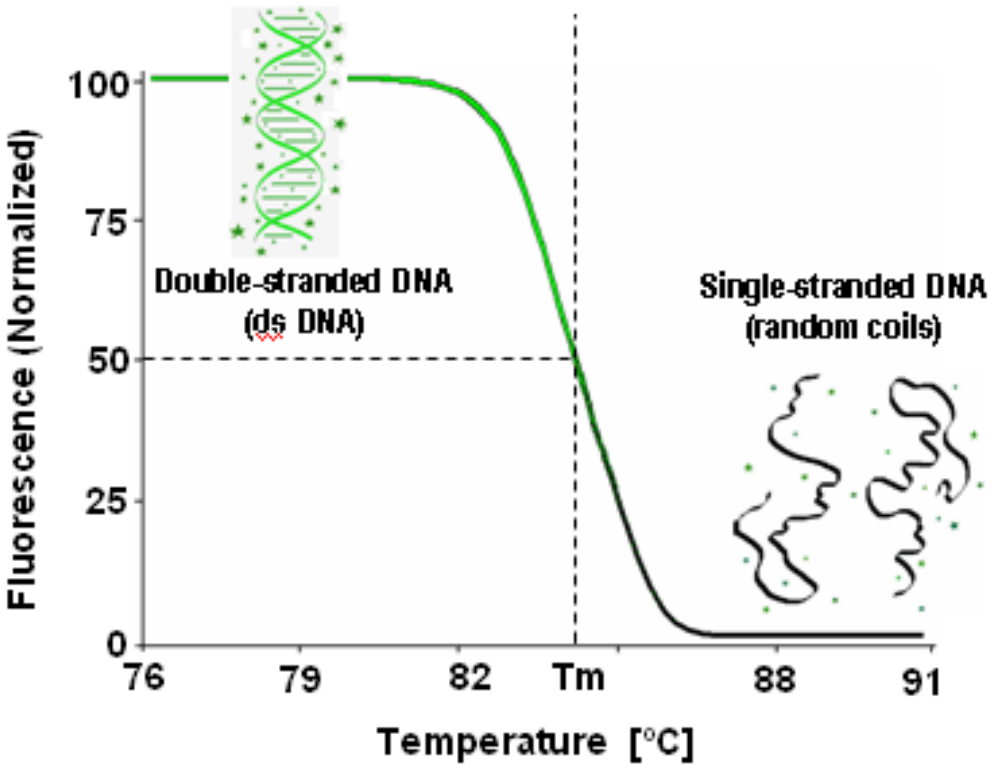

Figure 1: Double-stranded DNA opens during heat up and loses fluorescence and finally become single-strand DNA

Derivative was taken from the fluorescent dropping curve to give a fluorescent dropping speed curve. This curve is called high resolution melting curve. Usually a main peak will be seen in this curve indicating the point when the double stranded DNA completely open. The temperature of this point is called Tm. The most interesting part about this technology is that different sequence of DNA will give a different melting curve usually with different Tm, different shape of the main peak and sometimes small peak before the main peak. This is because various combination of nucleotide and the structure of DNA require different energy to open. Usually high GC content requires higher energy compared to high AT content, and compiled GC base pairs requires even higher energy to open. For example, for two sequences having the same length, the high GC sequence will have a higher Tm value. Therefore, we can use the feature of the melting curve to characterize difference sequence and identify different bacteria species.

1.2.3 Bacteria genus identification by HRM

Previous work has demonstrated the success of using the gene of 16 small ribosome RNA (16S rRNA) to detect bacteria pathogen by high resolution melt analysis (13). This gene has a length of 1542 base pairs (in E Coli) and includes 9 variable regions flanking by conserved regions. It is used for high resolution melting analysis for several reasons.

First, this gene is a conserved gene in all of the bacteria. Second, the variable region of this gene is different in most bacteria that provide the genetic difference for us to distinguish different bacteria species. Third, the 9 variable regions has already been widely studied and used for this purpose. Throughout the 9 variable regions, variable region V2 and V3 are most related to genus identification while region V6 is more related to serotype identification (14). In previous study, it shows that use universal primer for region V1 to V6 give us most information on the melting curve for identifying different genus of bacteria (15). Therefore, attempt already made to use this gene in high resolution melt analysis to differentiate different serotypes of Salmonella. However, through region V1 to V6 of different serotypes of Salmonella, there is only about 10 nucleotides difference which makes their melting curves look quite similar.

1.2.4 Digital PCR

Digital PCR is a new approach for DNA detection and quantification. Compared to the semi-quantification method qPCR, digital PCR can enumerate the target sequence at low concentration level. This test conducted on one digital PCR chip which consists 20,000 wells and enables such many picoliter-scale reaction happening simultaneously. Obeying the Pisson distribution, the concentration of DNA template loaded was carefully calculated to guarantee that in each well there will be one or no molecule. Like qPCR,

digital PCR also uses fluorescent reporter for positive amplification. Only well with target molecule can give out fluorescence. Therefore, absolute quantification of molecules can be visualized and calculated according to the number of fluorescent wells.

1.2.5 Massively parallel digital high resolution melt

Base on these two approaches, HRM and digital PCR, our lab developed a new technology, parallel digital high resolution melt. This technology performs HRM on the digital PCR chip under the fluorescent microscope following PCR reaction on the regular thermal cycler, and gives the melting curve of each amplicon within all the 20,000 wells. Using universal primer, this new digital high resolution melt platform technology enables the identification and absolute quantification of thousands of individual DNA molecule in a mixture (16). With this technology, we will know first how many molecules in the sample, second what the molecule is based on the feature of its melting curve and third find unknown sequence. We are trying to apply this new digital HRM technology to solve the Salmonella serotyping to see whether we can successfully identify different serotypes of Salmonella.

1.3 Salmonella and Salmonellosis

Salmonella is a rod-shaped gram-negative bacteria which is named after Dr. Salmon to memorize his founding of this bacteria. The bacterial genus Salmonella is divided into two species enterica and bongori. S.Enterica is further classified to 6 subspecies I, II, IIIa, IIIb, IV and VI according to two structures on the bacteria surface: surface covering-O antigen and part of the flagella-H antigen. There are more than 2500 serotypes within these two species and they are 95-99% identical at gene level (17). Among these serotypes, less than 100 serotypes from S.Enterica are associated with disease in human and domestic animal and about 50% cases of human infections can attribute to 3 serotypes: Typhimurium, Enteritidis and Newport (18,19).

Salmonellosis is a major cause of bacterial enteric illness. The food poisoning caused by Salmonella infection is called Salmonellosis. The bacteria usually stay in the intestine and cause diarrhea, vomit and stomachache. In most cases, the condition is not severe so the treatment is to provide rehydration and electrolyte, which can be achieved either by oral rehydration or intravenous solution. In severe case, antibiotic will be prescribed to stop the bacteria get into the blood stream and also help to keep the infection away from the baby for the pregnant woman. Among all the serotypes, Choleraesuis is most likely to spread to the blood to cause bacteremia (21).

Meanwhile, Salmonella can also cause other illness such as septicemic disease, typhoid fever, paratyphoid fever and in some rare cases cause osteoarticular infections (22).

1.4 Serotype identification for Salmonella

In Salmonella Enterica species, there are several serotypes that are most common to cause infection in humans. They are much similar in both genotype and phenotype and therefore cause much difficulty for identification. We therefore select them as the target for our studies including: Newport, Typhimurium, Heidelberg (UC1), Enteritidis (UC10) and Choleraesuis. As mentioned above, amplicon of V1 to V6 provides much feature in terms of shape diversity for genus differentiation, we will keep universal primer for region V1 to V6 in our design but with the intention to advance the HRM analysis for serotype identification. The difficulty for serotype identification relies on the few sequence difference among them. Through the region V1 to V7, there are only 12 nucleotides difference among all the 5 serotypes (NCBI database), and 8 different loci are in the V6 and V7 region. So whether we can tell this few different nucleotides out of the ~1000 bp long sequence becomes our biggest question. For this purpose, two approaches were investigated. The first is preferential intercalating dyes, and the second is multiplex PCR.

1.5 Data analysis by machine learning

Support vector machines (SVM) is a supervised learning models that analyzes data used for classification and regression analysis and was proved to be an efficient method to classify data from HRM. SVM will use part of the data for training and build a model, and then test the data left to assign them to one category or the other. We will use this algorithm to test whether the melting curve of different serotypes are distinguishable.

1.6 Antibiotic resistance

1.6.1 Antibiotic resistance as a global problem

Today, antibiotic resistance is one of the most pressing problems in public health and will remain threatening the modern medicine in the coming decades. Because of the increasing abuse of antibiotic in hospitals and community, some widely used antibiotics are losing their function and scientist and doctors must develop new antibiotic to overcome this problem. The arrival of untreatable strains of carbapenemresistant Enterobacteriaceae indicates we are at the dawn of a postantibiotic era (23). Many efforts have been made to address this problem including research and new standard in antibiotic prescription and etc.. Last year the National Institute of Allergy and Infectious Diseases (NIAID), part of the National Institutes of Health, has awarded more than $11 million in

first-year funding for nine research projects supporting enhanced diagnostics to rapidly detect antimicrobial-resistant bacteria (24). However, more attention should be paid to this global problem especially the coordinated action at the political level both nationally and internationally (25).

1.6.2 How do bacteria become antibiotic resistance

Some bacteria are naturally resistant to some antibiotics, for example most organisms found in human gut are resistant to benzyl penicillin (26). Some bacteria can develop resistance to the antibiotic that was usually used to kill it. Generally, bacteria can develop antibiotic resistance in two ways: by genetic mutation or by acquiring resistance from another bacterium (27).

Rare mutation happens in the bacteria genomes. Some mutations enable the bacteria to produce enzyme to inactivate the antibiotics and other mutations can eliminate the antibiotics target in the cell. Even the mutation is rare, occurring in one in one million to one in ten million cells, the large number of the bacteria and its rapid reproduction rate makes this a possible approach to acquire antibiotic resistance. Especially in the clinical settings where the antibiotic selection pressure is much higher, following Darwin's evolution theory, it is even easier for the antibiotic resistant bacteria to survive. For the

same reason, unnecessary prescribing and overprescribing in hospital and community, easiness to buy and use for general customers and unrestricted use in agriculture aggravate this problem. This problem makes it harder for doctors and healthcare professionals to choose the effective antibiotic for their patients and sometimes doctors even run out of choice for their patients with infections.

Meanwhile, bacteria can acquire antibiotic resistance from other bacteria. Genetic material such as gene encoding resistance to antibiotic found on plasmid and transponos can be transferred through a simple mating process called “conjugation” (28). This exchange of gene can even occur between different bacterial species which makes the spreading of antibiotic resistance more convenient.

1.6.3 Antibiotic detection in hospitals

In hospitals, bacteria antibiotic resistance is detected by culture the bacteria in the medium with different type of antibiotics. This takes at least one day and can give false negative or positive result. Meanwhile, the antibiotic resistance might not come from the bacteria itself but from the contamination or the environment that come with the bacteria sample for which the current method cannot tell.

Therefore, we try to extend the application of the multiplex PCR and digital HRM to this area to help detect the antibiotic resistance in different bacteria. We will use multiplex PCR to amplify both the bacteria 16S rRNA gene and antibiotic resistance gene and analyze its melting curve. As digital PCR only allow at most one molecule to presence in the well, we will be able to know what this molecule (bacteria) is and whether this molecule has antibiotic resistance. This will not only shorten the detection time under 3hours, but also complete the bacteria identification and antibiotic resistance detection in one test.

CHAPTER 2: Serotype identification by different intercalating dyes

2.1 Introduction: intercalating dye

Intercalating dye is a type of fluorophore. It can specifically bind to double-stranded DNA and give out fluorescence. In the absence of double-stranded DNA, it binds to nothing so the fluorescence is low. There are different family of intercalating dyes classified dependent on their structures and original sources. The most commonly used dye includes Eva green and SYBR green. In recent years, Eva green is more widely used in the research laboratories because of its non-selective binding and no inhibition on PCR. While SYBR green has shortcomings including preferential binding to G/C base, inhibition to PCR reaction and low reproducibility. However, we want to make use of its preferential binding capability to help us tell the difference subtype of Salmonella.

Among all the intercalating dyes, TOTO family of dyes are among the most sensitive and highest affinity fluorescent dyes available for nucleic acid staining (29). They are symmetric dimeric nucleic acid stains, including TOTO-1,TOTO-3, BOBO-1, BOBO-3, POPO-1,POPO-3,YOYO-1, YOYO-3, JOJO-1 and LOLO-1. Among these TOTO-3 is reported to have preferential binding to 5'-'CTAG' -3' sequence, and BOBO-3 to AT bases.

2.2 Experiment Design

Our design relies on the hypothesis that the preferential binding will magnify the nucleotide difference through the melting curve. We think that with the help of G/C preferential binding dye, the sequence has more G/C base pairs will has its Tm shift to right or a higher peak compared to the sequence has less G/C pairs. For the same reason, A/T preferential binding dye and other dyes that has specific preferential sequence to bind will help to magnify the difference among the serotypes.

With this in mind, we select BOBO-3, TOTO-1 SYBR Green as long as the control dye EVA Green as the fluorescent reporter for our studies. They were added for qPCR followed by melting curve test and data analysis. We tested whether these preferential binding dyes can magnify the difference of sequence to the extent that the a few base pairs difference within the serotypes can be distinguished by their melting curves. Because SYBR Green, BOBO-3, and TOTO-1 were reported to inhibit amplification in high concentration, we conducted experiment by adding them both before and after PCR reaction. Each test has 6 replicates to test the reproducibility.

Different dye concentration was also studied to see at which concentration we can obtain clear fluorescent signal and minimize noise. Different melting speeds were tested

to see whether this would influence the result.

Data were analyzed by eye and support vector machine (SVM) algorithm. Raw fluorescent derivative data was generated by the software and used for analyzing. In SVM, curve of 5 replicates of each serotype were taken out randomly as training curves, and the left curve was tested to see which serotype it belongs. We run the random training curve picking process 20 times, and calculate how many times one serotype was classified as another serotype.

2.3 Method and Material

2.3.1 Bacterial genomic DNA Isolation

Five serotypes of Salmonella that are selected: Newport, Typhimurium, Heidelberg (HEIDELBERG), Enteritidies (HEIDELBERG0) and Choleraesuis. They were cultured in the LB agar plate overnight. Several colonies were picked and cultured in a 5ml Luria-Bertani (LB) broth overnight. Then we used the Wizard Genomic DNA Purification Kit (Promega) to isolate genomic DNA (gDNA) from the overnight medium culture of bacteria. Spectrophotometer reading was taken to assese the quality and quantity of the extracted gDNA.

2.3.2 PCR efficiency

The efficiency of PCR can be evaluated by performing a dilution series experiment using the target sample. We used a 10-fold dilution series, diluting the extracted DNA by 10, 100 and 1000 times. The slope of the standard curve was translated into an efficiency value. 10 fold concentration was finally used for qPCR.

2.3.3 Primer design

Region V1 forward primer and V6 reverse primer are from paper (19).

V1F 5'-GYGGCGNACGGGTGAGTAA-3'

V6R 5'-AGCTGACGACANCCATGCA-3'

2.3.4 qPCR and High Resolution Melt

Reactions were performed in a 15 ul final volume and contained following final concentration: 1X PCR buffer (Phusion HotStart), 0.15 uM forward primer, 0.15 uM reverse primer, 0.2 mM dNTPs (Invitrogen, Carlsbad, CA), 1X EvaGreen (Biotium, Hayward,CA) for control reactions, 0.15 uM dNTPs (), 0.2 U/ul Phusion HotStart polymerase, 1.5 uL of sample, and ultra pure PCR water (Quality Biological, Gaithersburg, MD) to bring the total volume to 15 uL. qPCR reaction was performed on

the ABiolife SDS3.2 following the cycle: an initial enzyme activation (98 degrees C, 30 s), followed by 35 cycles (98 degrees C, 10 s, 59 degrees C, 30 s, 72 degrees C, 60 s), and an elongation step (72 degrees C, 5min).

For qPCR reaction with other 3 dyes: SYBR Green, BOBO-3 and TOTO-1, SYBR Green was added before the reaction at suggested 1X to see whether it will inhibit amplification. A final concentration of 0.1 uM of BOBO-3 or TOTO-1 was tested by adding them before PCR. A final concentration of 0.625 uM of BOBO-3 or TOTO-1 was tested by adding them after PCR.

Melting curve was obtained following the melting steps: DNA double strands opening step - 95 degrees C, 15s, gradually heating step - 45 degrees C, 15s to 95 degrees C with a temperature increase of 2% (suggested optimum speed) of the maximum setting of the machine. A ramping rate of 1% was also tested to see whether slower speed helps show more information about the melting curve.

2.4 Result

2.4.1 DNA extraction

DNA quality was evaluated by its concentration, DNA to protein ratio (260nm/280nm) and DNA to RNA ratio (260nm/230nm) shown in Table 1. DNA with 260/280 and 260/230 ratio at the range of 1.7-2.0 can be used for amplification.

Table 1: DNA quality

Serotypes	Newport	Typhimurium	Heidelberg	Enteritidis	Choleraesuis
Concentration ng/ul	220	180	256	190	278
Absobtion 260/280	1.91	1.94	1.93	1.84	1.9
Absobtion 260/230	1.78	1.9	2.14	1.87	1.94

2.4.2 PCR efficiency

Ct value of different dilution was shown in Table 2 and Figure 2. The theoretical difference of Ct value for every 10 fold dilution should be 3.3.

Table 2: PCR efficiency

Dilution factor	1	10	100	1000
Newport	9.0	11.1	15.8	20.1
Typhimurium	10.7	13.4	17.2	22.5
Heidelberg	8.9	11.9	16.4	20.4
Enteritidis	8.2	10.3	14.1	18.5
Choleraesuis	8.0	10.4	14.9	18.9
NTC		27.7		

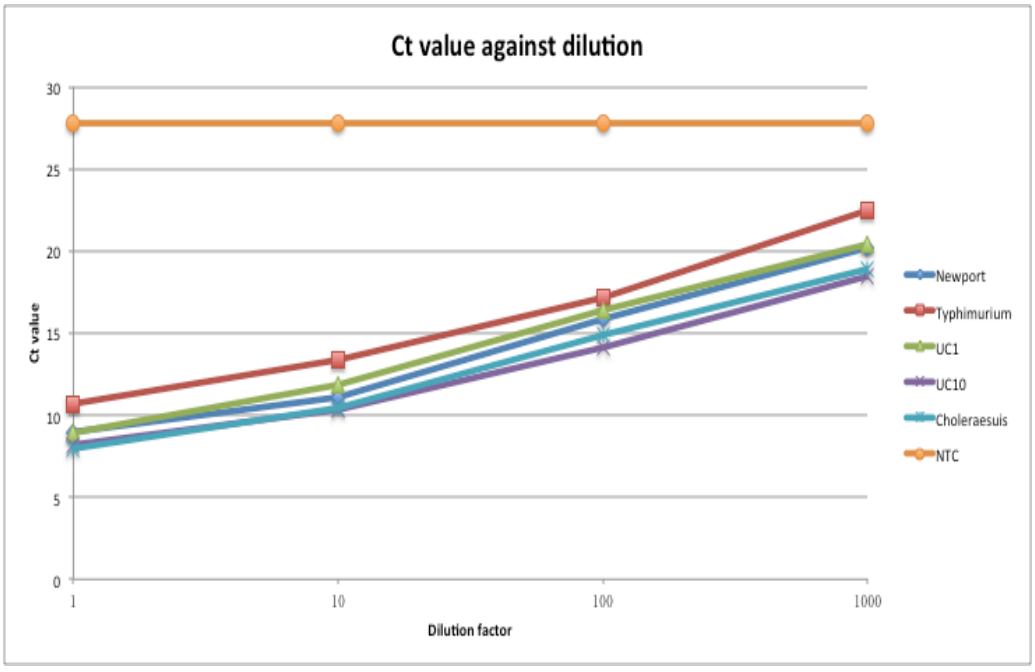

Figure 2: PCR efficiency in 10-fold dilution

2.4.3 Melting curve with 4 intercalating dyes

Melting curve with Eva shown in Figure 3: Small difference is shown between Heidelberg and other four serotypes. Tm is around 90.5C. So Eva can't distinguish the five serotypes.

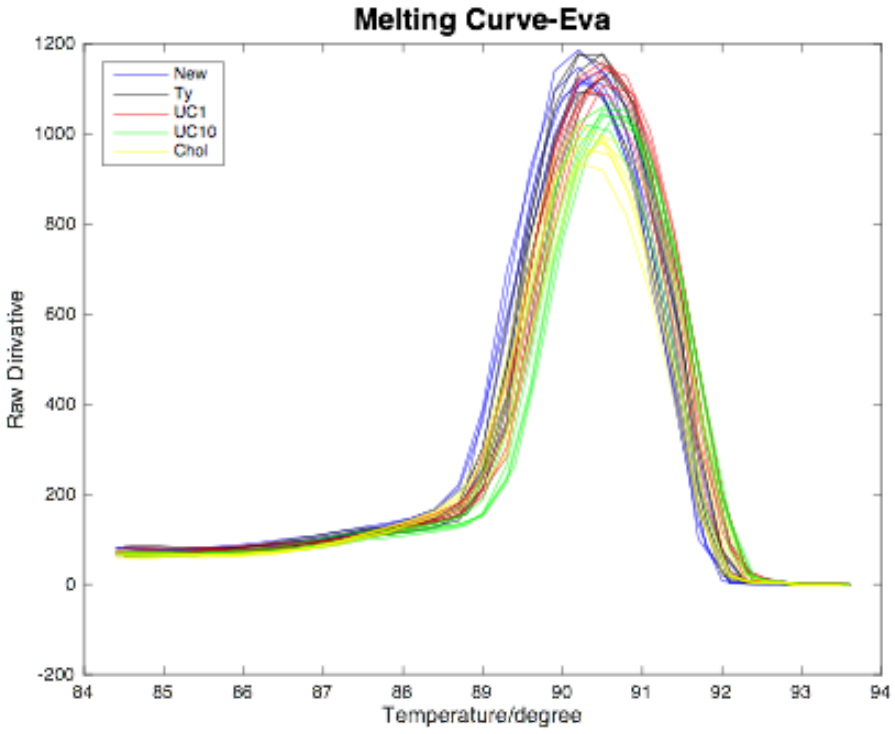

Figure 3: Melting curves of 5 serotypes with Eva Green

Melting curve with SYBR Green shown in Figure 4: No difference is shown in the five serotypes. So SYBR can't distinguish the five serotypes.

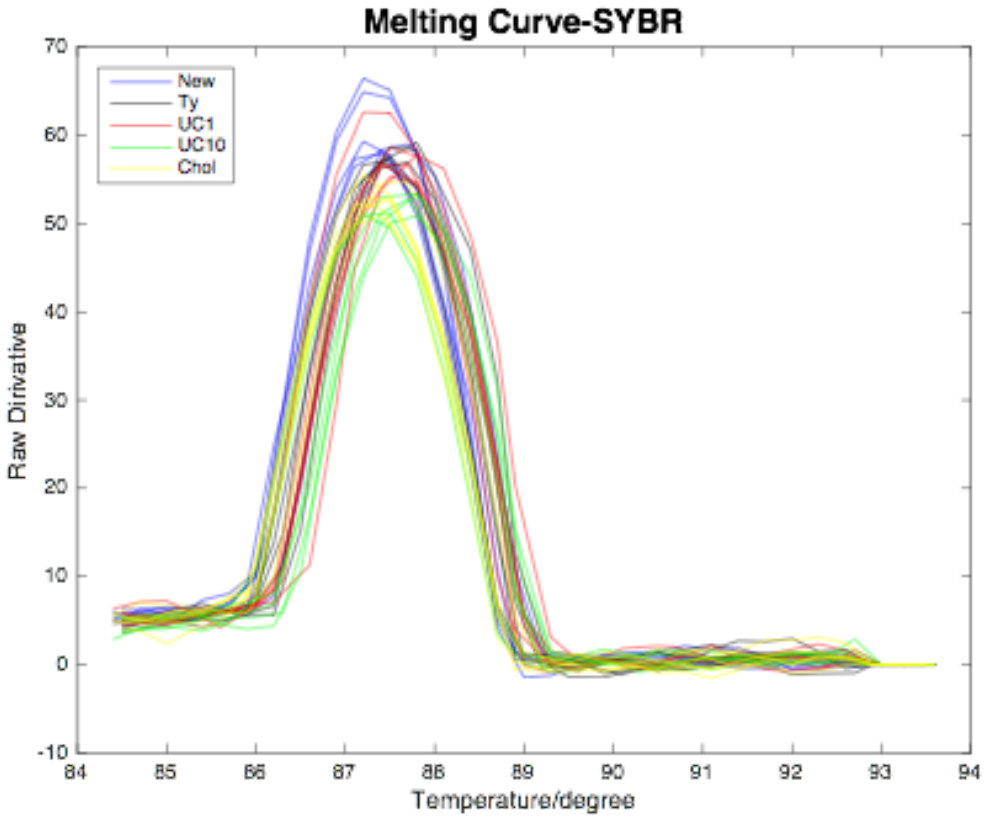

Figure 4: Melting curves of 5 serotypes with SYBR Green

Melting Curve with BOBO-3: BOBO-3 added before PCR completely inhibit the amplification even at a final concentration of 0.1uM. Therefore, we added the dye after PCR and found at a final concentration of 1.67uM, the dye can give clear signal and relative small noise shown in Figure 5. Like Eva Green, HEIDELBERG0 is distinguishable from the cluster of curves. Meanwhile, Newport and Typhimurium seems to be categorized as one group and Heidelberg and Choleraesuis as another group.

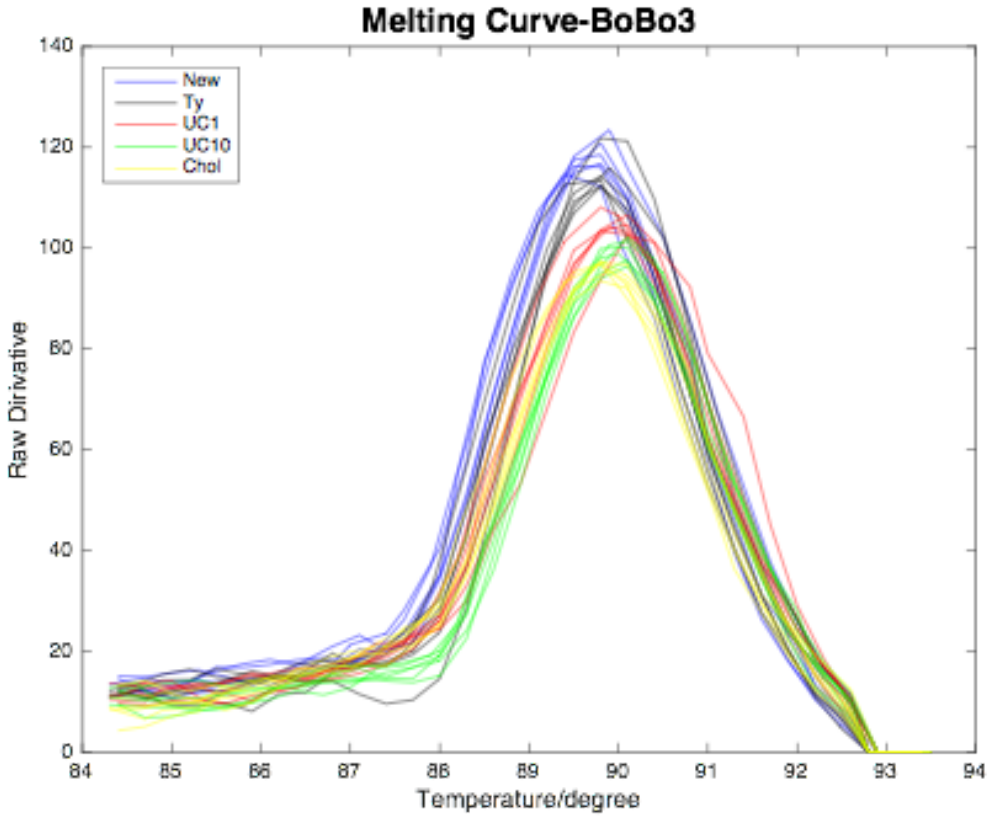

Figure 5: Melting curves of 5 serotypes with BOBO-3

Melting curve with TOTO-1: as BOBO-3, TOTO-1 inhibited PCR at 0.1uM, so was added after. At concentration 1.67uM, the dye can give clear signal and relative small noise shown in Figure 6. Similarly, it's hard to differentiate the five serotypes through their melting curves.

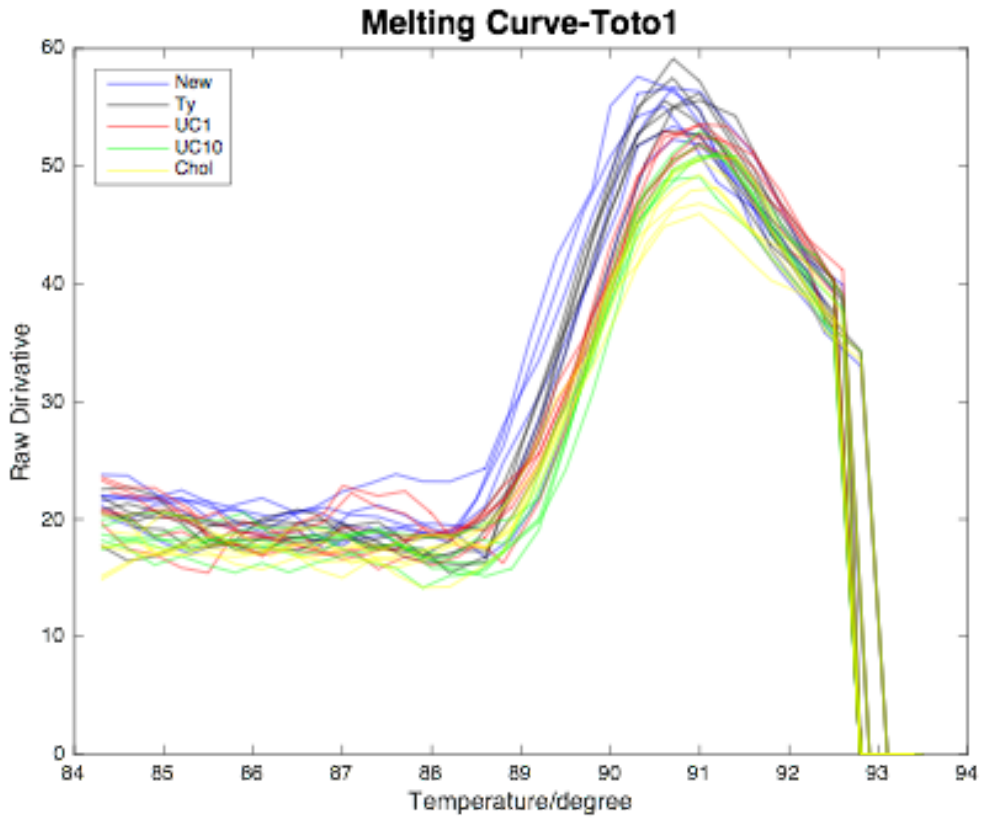

Figure 6: Melting curves of 5 serotypes with TOTO-1

2.4.4 Influence of temperature ramping speed

To evaluate whether slower ramping rate will help to generate more feature on the melting curve, 1% ramping rate was tested on BOBO-3 and TOTO-1 shown on the right side of Figure 7. Lower fluorescent derivative and higher noise were seen without any improvement for differentiating different group of curves. Therefore, we didn't further investigate other ramping speed and use the result of the recommended 2%.

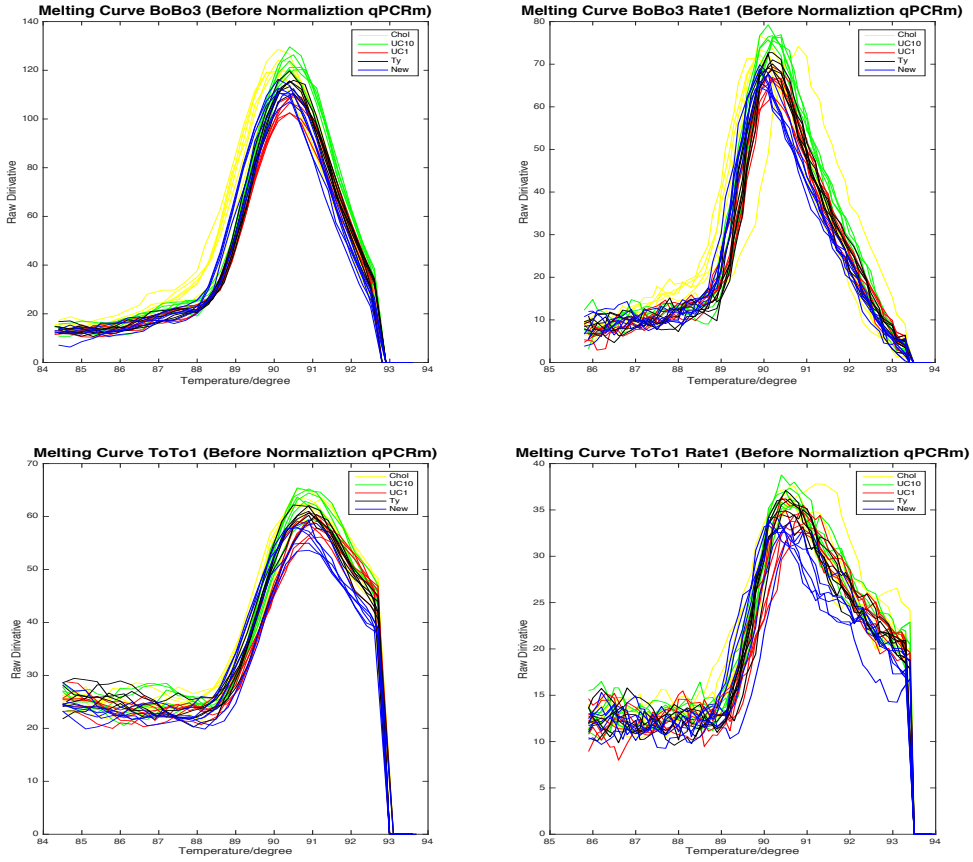

Figure 7: Melting curves of BOBO-3 and TOTO-1 at 1% and 2% ramping speed

2.4.5 Data analysis by machine learning

We use SVM algorithm to analyze the result above to see whether machine learning can better classify the melting cure of each serotype. Error map as shown in Table 2-6 lists within 20 times of testing how many times one serotype was classified as another serotype, e.g. curve of Newport was recognized as Typhimurium for 1 time and as Heidelberg for 6 times.

Table 3: Error map of Eva Green

Error	As New	As Typh	As Heid	As Enter	As Chol
New	0	1	6	0	0
Typh	3	0	6	2	0
Heid	0	3	0	3	0
Enter	0	9	1	0	0
Chol	0	0	0	0	0

Table 4: Error map of SYBR Green

Error	As New	As Typh	As Heid	As Enter	As Chol
New	0	0	10	0	0
Typh	2	0	1	0	3
Heid	8	4	0	5	3
Enter	0	7	0	0	10
Chol	0	3	0	14	0

Table 5: Error map of BOBO-3

Error	As New	As Typh	As Heid	As Enter	As Chol
New	0	0	0	0	0
Typh	6	0	0	0	0
Heid	0	0	0	5	0
Enter	0	0	6	0	0
Chol	0	0	0	0	0

Table 6: Error map of TOTO-1

Error	As New	As Typh	As Heid	As Enter	As Chol
New	0	9	2	0	0
Typh	6	0	2	0	0
Heid	0	0	0	3	0
Enter	0	0	5	0	8
Chol	0	0	0	6	0

Result from SVM gives the same conclusion as we obtain by eye before: EVA Green can reproducibly identify Choleraesuis; SYBR can't identify any serotype; BOBO-3 can also identify Choleraesuis, but misidentify Typhimurium to Newport, and Heidelberg to Enteritidis; TOTO-1 can't identify any serotype.

CHAPTER 3 Multiplex PCR

3.1 Introduction

Multiplex PCR is to use more than one pair of primer to amplify several DNA fragments in one reaction. For bacteria identification, gene of 16S rRNA region is widely studied for bacteria classification. The whole length of this gene is 1452 bp (in E coli.), including 9 variable regions flanking by conserved regions. Among all the variable regions, region V2 and V3 are most suitable for distinguishing bacteria to the genus level, V6 to the species level and V4, V5, V7 and V8 are less useful for bacteria identification (20). In previous studies, regionV1 to V6 is proved to be efficient for species identification and this long fragment shows diversity in shape which can even help to tell one nucleotide difference (21). However, as amplify V1 to V6 region only is difficult to identify Salmonella to the serotype level, we tried to include more region for this purpose.

3.2 Design

Sequence difference between the 5 serotypes was analyzed. V1 to V6 region (V16) with length about 1000bp was selected for bacteria species identification. Then we want to pick a shorter region where most different base pairs were show within the 5 serotypes to see whether the curve of this fragment can distinguish the serotypes. There are a total 12 base pairs of difference through this gene, 4 in V1 to V5 region, 4 in region V6 (~58 bp), and 4 in region V7(~57 bp). V6 and V6+V7 (V67) two fragments were therefore picked because they consist most different base pairs. Primers were then designed for covering the whole V6 region and V6+V7 region respectively.

Thinking amplicon with different length will bind different amount of dye and give out fluorescence of different intensity, we are concerned peaks of these two fragment V16 (~1000bp) and V6 (~125bp) will be biased. Therefore, primer ratio of V16 to V6 region were tested to guarantee that fluorescent signal from both fragments can be seen. Forward primer ratio of long amplicon to short amplicon tested were 2, 1, 1/5 and 1/10.

3.3 Method and material

DNA isolation, qPCR and high resolution melt were the same as in the chapter 2 with the only difference of primer. Total primer concentration for both forward and reverse primer is kept at 0.15 uM with forward primer splitting to two primers: V1 forward primer and V6 forward primer.

V6 forward primer V6-F (5'-GGAGCATGTGGTTTAATTCGA-3') is from another paper published by Dr. Fraley (19). The reverse primer for V6 is the same reverse primer for V16 region. This pair of primers gives a 129bp long amplicon for V6 region.

Reverse primer for V7 region was designed by NCBI Primer Blast: 5'-CTTGACGTCATCCCCACCTT-3' and pairing with the forward V6 primer, gives a 254bp amplicon.

3.4 Result

3.4.1 Melting peak of most variable region V6 and V6+V7

Melting curve of V6 region show in Figure 8: this region can identify 3 serotypes: Newport, Typimurium and HEIDELBERG by displaying clear difference in the "pre-peak" area, but fail to distinguish between HEIDELBERG0 and Choleraesuis. Tm of main peaks is around 86C.

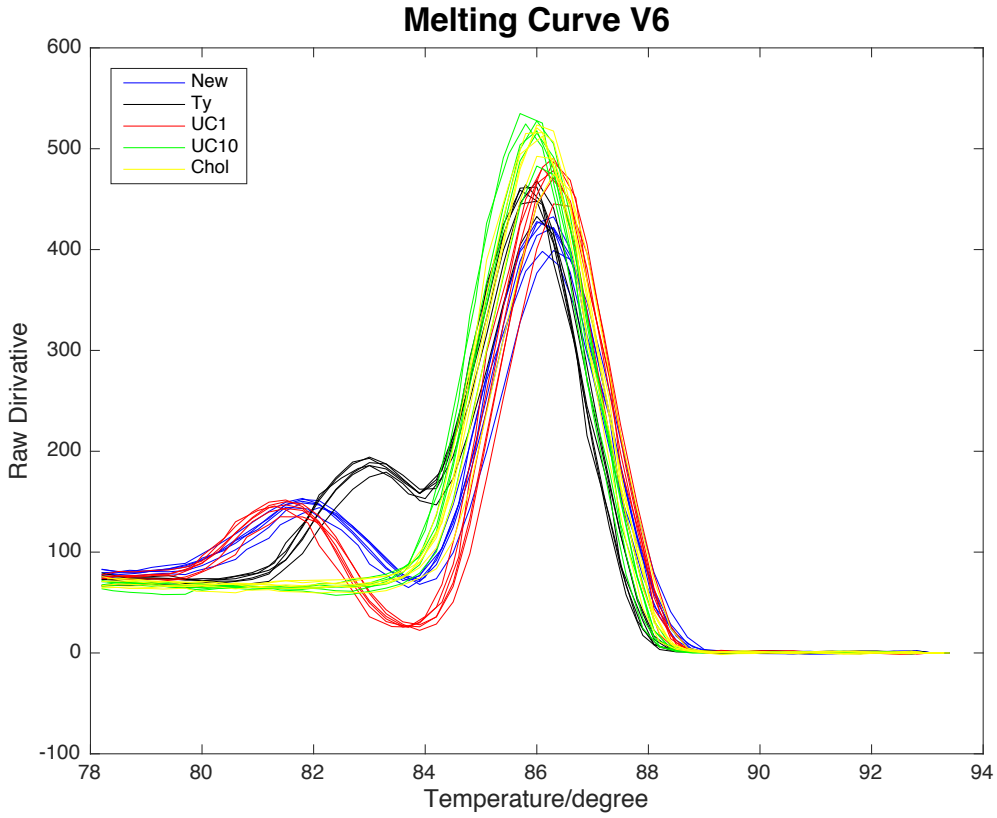

Figure 8: Melting curve of V6 region for 5 serotypes

Melting curve of V67 shown in Figure 9: This region can tell all the 5 serotypes apart by showing difference in the per-peak area, peak height and Tm.

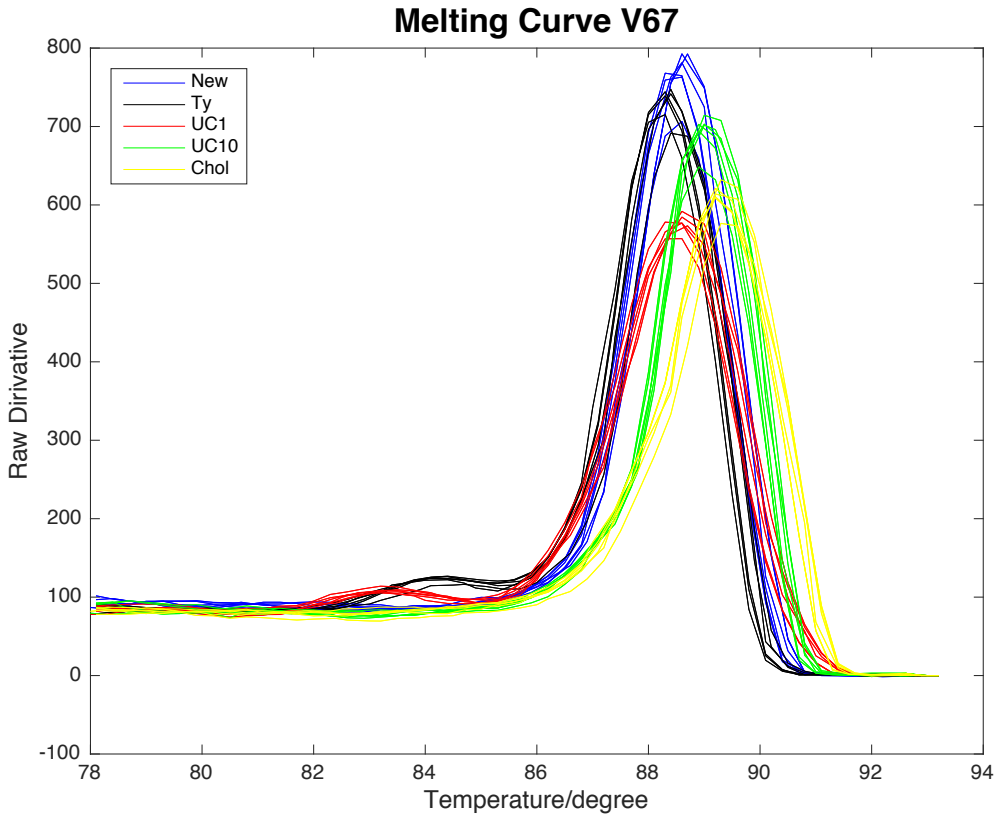

Figure 9: Melting curves of V67 region for the 5 serotypes

However, the difference in the pre-peak area is not as obvious as from the V6 region only. Also, because of the longer length of this region (~254bp) compared to V6 (~129bp), the Tm falls in the range of 88C to 90C. As we want to use this region combined with V16 region (peak Tm ~91C), and make sure both peaks shows up with minimum overlapping. We decide to use V6 to complement V16 region.

3.4.2 Primer ratio

Result of different primer ratio of V16 to V6 (Typhimurium) was shown in Figure 10: melting peak of V6 didn't show up when there are equal or more primer of V16. This can be explained that the long amplicon binds more fluorescence and hallow the peak of the short amplicon. Therefore, a high primer ratio for short amplicon is needed. At primer ratio 1:5 and 1:10 of V16 to V6, two peaks are clearly seen at the point where the individual peak is previously showed up. Therefore we picked these two ratios to test all the 5 serotypes.

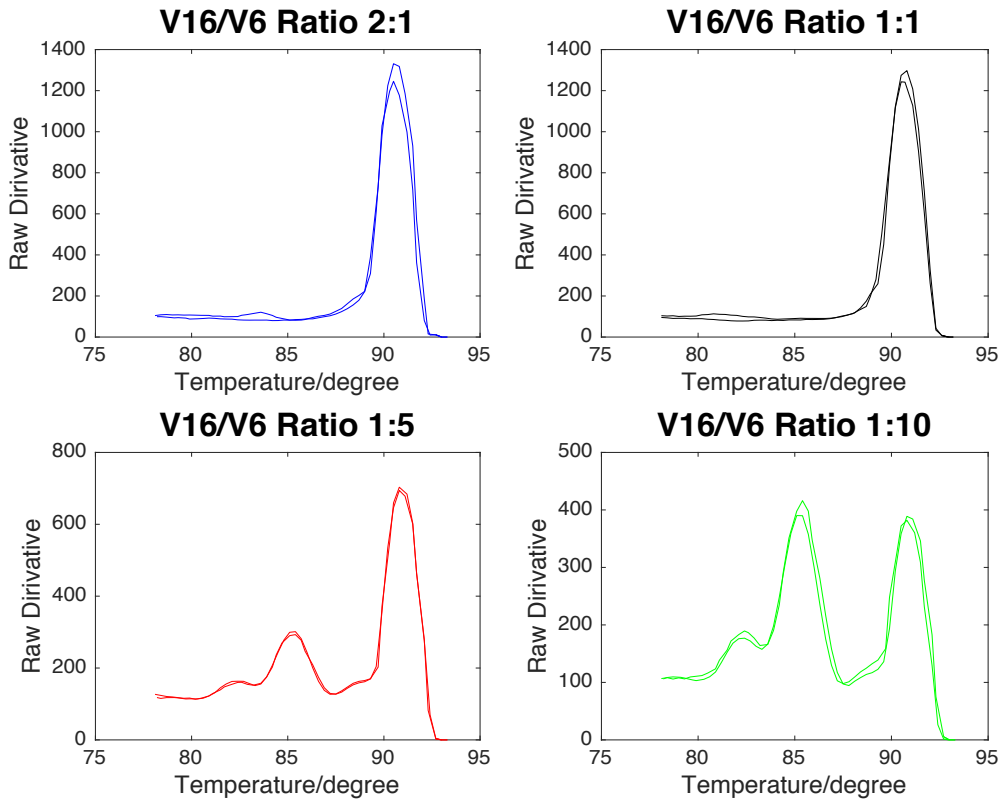

Figure 10: Melting curve of V16 and V6 region of Typhimurium with different primer ratio

3.4.3 Melting curve of 5 serotypes with V16 and V6 region

Melting curve of V16+V6 with primer ratio 1:5 of V16 to V6 shown in Figure 10: this set of curves gives a clear distinction for the 5 serotypes, specifically the small peak of V6 region helps identify Newport, Typhimurium and HEIDELBERG, and main peak of V16 region helps distinguish between HEIDELBERG0 and Choleraesuis.

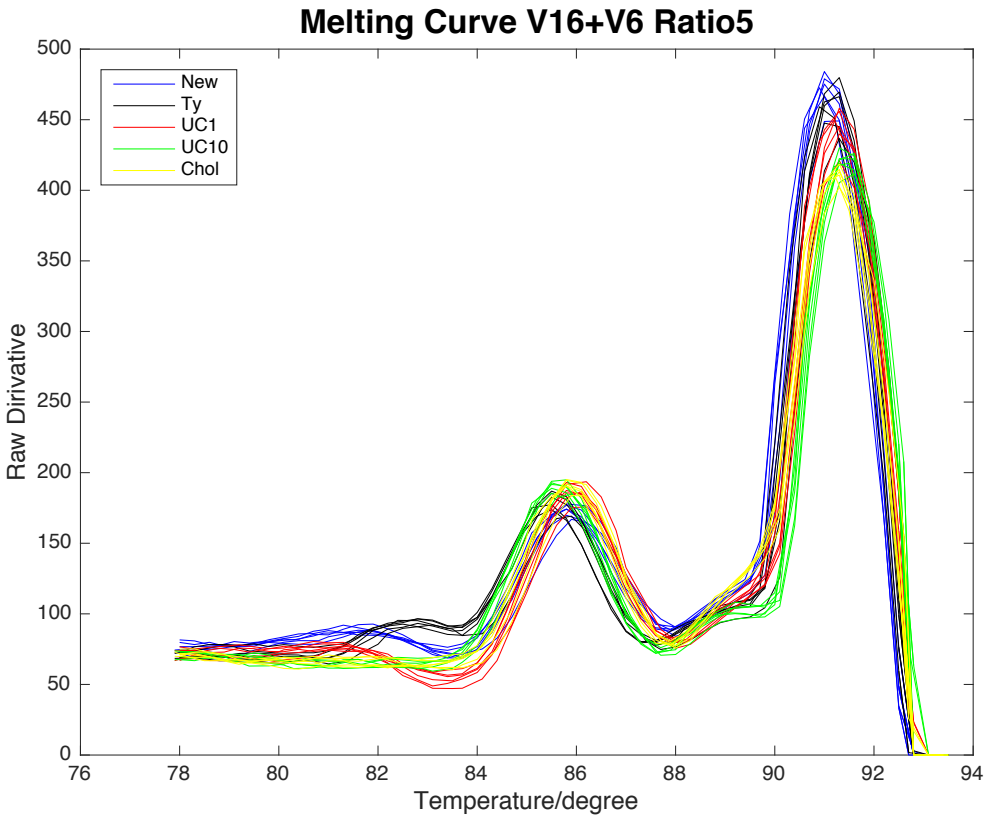

Figure 11: Melting curves with primer ratio of the V16 to V6 at 1:5

Melting curve of V16+V6 with primer ratio 1:10 of V16 to V6 shown in Figure 11: this set of curves also gives a clear distinction for the 5 serotypes with a bigger peak in V6 region.

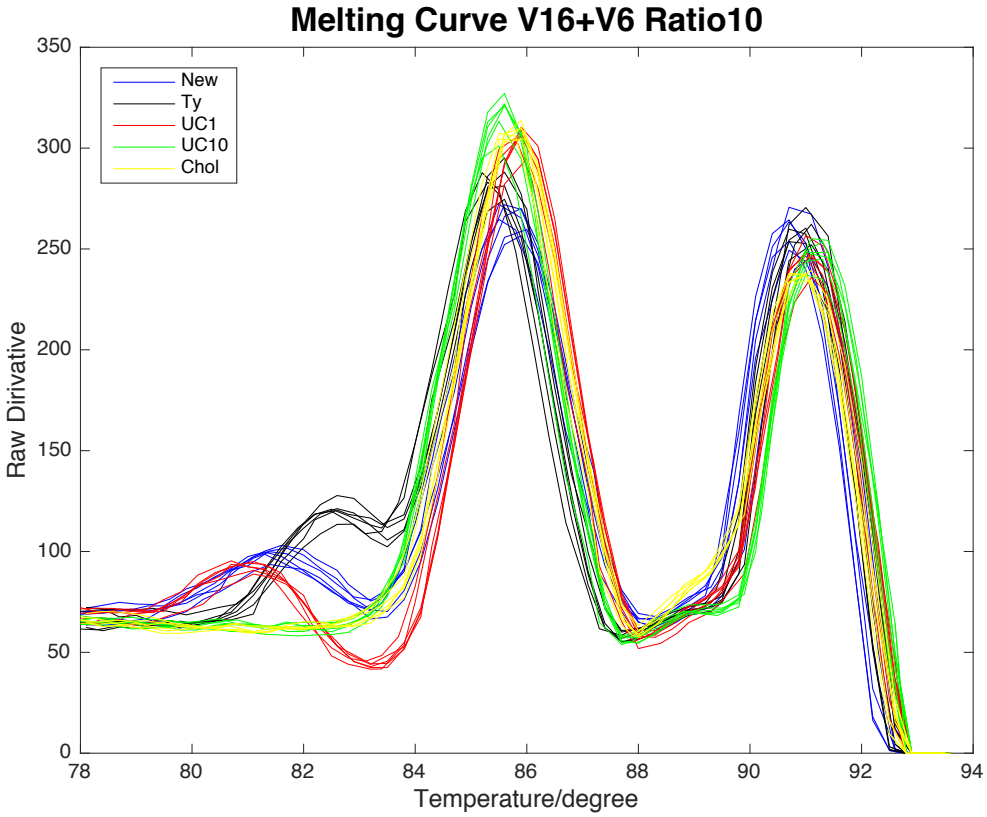

Figure 12: Melting curves with primer ratio of the V16 to V6 at 1:10

CHAPTER 4 Antibiotic Resistance

4.1 Introduction

Antibiotic resistance is a vital problem in health care and increasingly draw people's attention in these years. Besides bacteria which is resistant to only one type of antibiotics, there arises multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria (30). Because of antibiotic resistance, health workers and scientist must continuously develop new antibiotic and inform the public to not abuse antibiotic, otherwise the development of resistance will outpace our capability to provide new antibiotic and this will threat the modern medication globally. In hospital, many bacteria carries antibiotic resistant gene because of the antibiotic resistance transferring between species and the selection pressure. Therefore to develop an efficient and robust method to tell antibiotic resistant comes to our attention.

Following the idea we tried above, we want to develop a method that can realize bacteria identification and antibiotic detection in one test by multiplex PCR and HRM. Our ultimate goal is to design a series of primers, which can detect different antibiotics with distinguishable melting peak from the HRM and also a main peak which can identify the bacteria species.

4.2 Design

We started our trial with the E. coli with Kanamycin resistant gene artificially inserted to its genomic DNA. Because we want to have a distinguishable peak of kanamycin resistant gene away from the main peak for bacteria identification, we need to pick the region from the Kanamycin resistant gene that is short enough to give a lower Tm value but long enough to be amplified. Therefore we use Matlab to analyze the GC content of the whole gene. The lower the GC content, the lower the Tm value. Finally, the region with lowest GC content was picked. Primers were selected to give the shortest fragment to guarantee the lower Tm. Potential fragments were then analyzed by U-Melt to predict its Tm value in HRM analysis. For bacteria identification, we used the region V16 plus V6 with a primer ratio 1:5.

4.3 Method and Material

Wild type E coli K12 GM1655 and E coli inserted with Kanamycin in genome were kindly provided by Ye Gao in Dr. Palson's lab. These two bacteria were cultured and extracted using the same method as described in chapter 2. qPCR and high resolution melt were also performed the same way.

Region of 50 base pairs with lowest GC valued was scrutinized by Matlab and primers for this region were designed by NCBI Primer Blast:

F-5'-CTTGCCGAATATCATGGTGGAA-3'

R- 5'-GATGAATCCAGAAAAGCGGCCA-3'

The final length of the fragment is 46 bp with Tm at 76.5C predicted by U-Melt. Primer ratio of 5:5:1 of Kanamycin fragment to V6 to V16 was selected based on our previous experience.

Gel electrophoresis was use to confirm the amplification product.

4.4 Result

4.4.1 Gel electrophoresis

Result was shown in Figure 12: one correct target band was seen in the amplification of V16, V6 or Kanamycin fragment. 3 target bands: V16, V6 and Kanamycin were shown in E Coli with Kanamycin resistant, but no Kanamycin fragment shown in wild type E Coli. Untargeted band around 500bp was seen in all the amplification having V16 region, which we think is a misamplification.

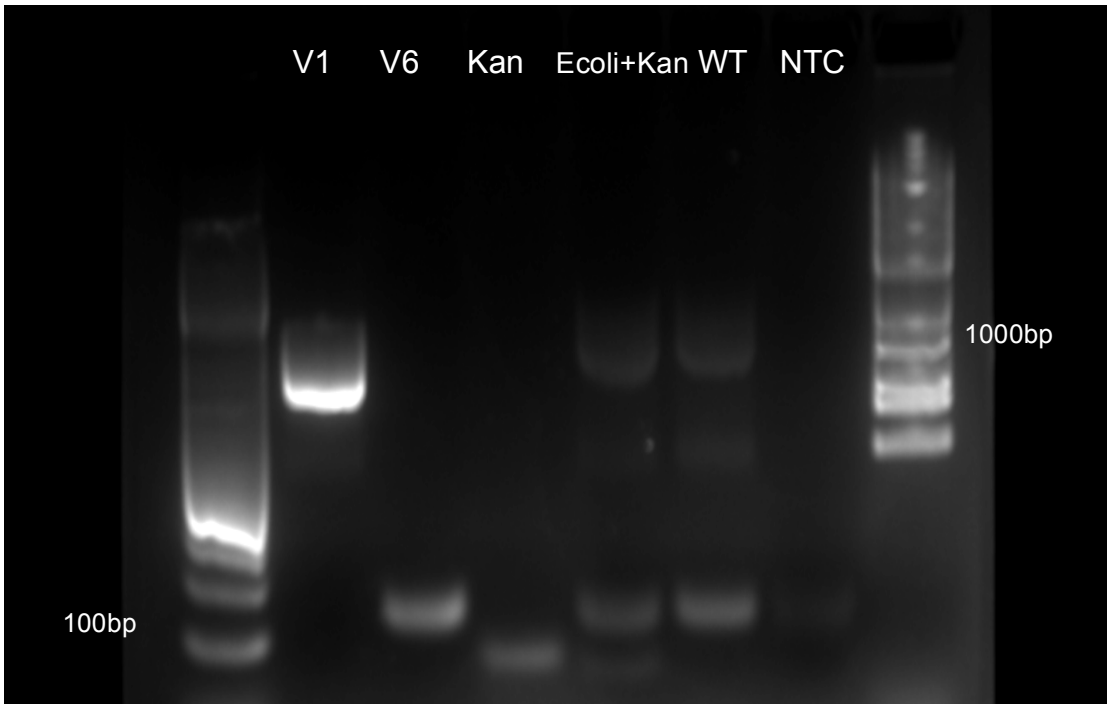

Figure 13: Gel electrophoresis of V1, V6, Kanamycin fragment, E Coli with Kanamycin resistance and wild type E Coli and NTC.

4.4.2 Melting curves

Melting curve of Kanamycin, V16 and V6 region is shown in Figure 13: V16 peak with Tm ~91C, V6 with Tm ~86C (a small peak at 81C is part of the V6) and Kanamycin peak with Tm~76C.

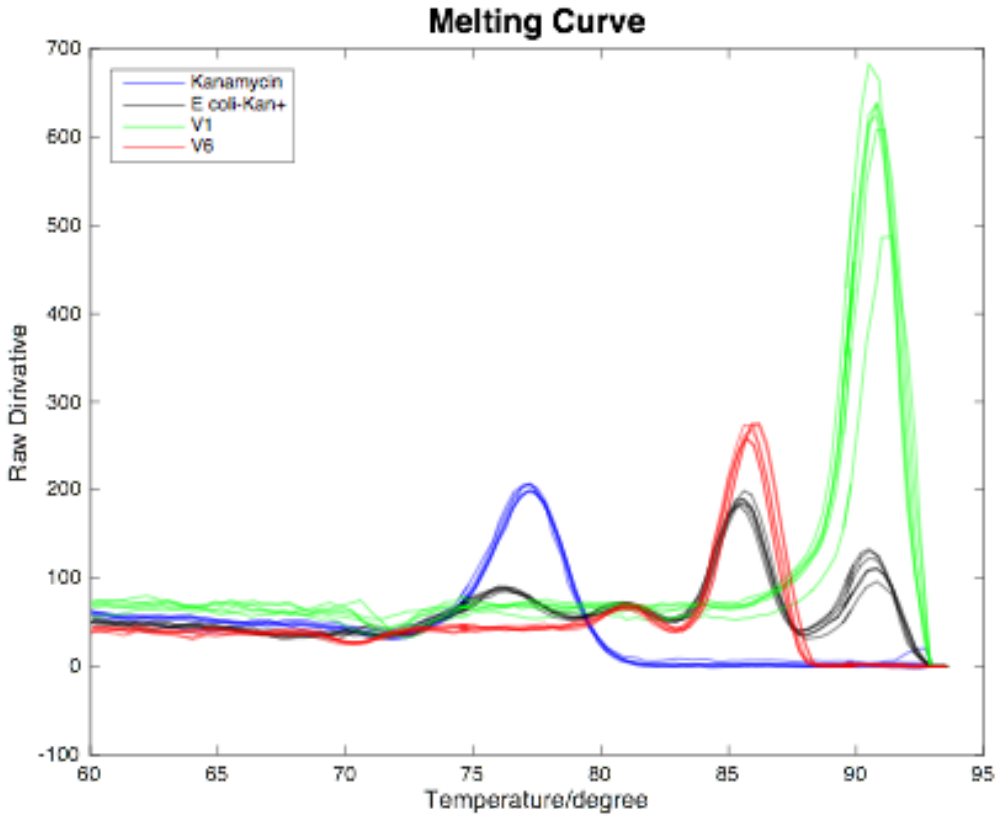

Figure 14: Melting curve of V16, V6, Kanamycin and E Coli with antibiotic resistance

Melting curve of 3 regions from E coli +/ Kanamycin and Wild type is shown in Figure 14: only the Kanamycin resistant type has the peak at Tm ~76C.

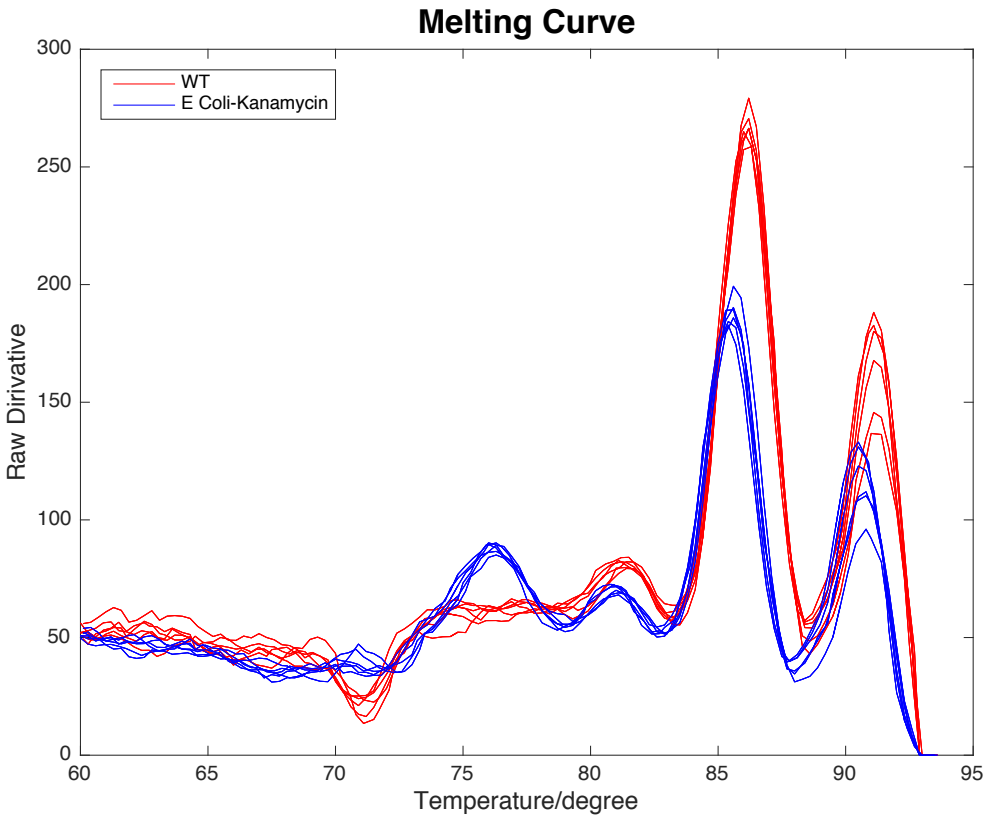

Figure 15: Melting curve of E Coli with antibiotic resistance and wild type

CHAPTER 5 Conclusion

5.1 Bacteria identification by intercalating dye

For four type of intercalating dyes investigated in this study: Eva Green, SYBR Green, BOBO-3 and TOTO-1, none of them can distinguish the 5 serotypes of Salmonella through HRM.

5.2 Bacteria identification by multiplex PCR

HRM following multiplex PCR of region V16 and V6 of 16S rRNA with primer ratio of V16 to V6 1:5 or 1:10 can help to distinguish all 5 serotypes of Salmonella.

5.3 Bacteria antibiotic resistance detection combining species identification

HRM following multiplex PCR of V16, V6 of 16S rRNA and fragment of Kanamycin resistant gene with primer ratio of V16 to V6 to Kanamycin 1:5:5 can help to detect both the bacteria species and antibiotic resistance.

5.4 Future research

Further work can be to investigate whether serotype identification by multiplex PCR of V16 and V6 region can be applied to other bacteria. For bacteria that have difference in gene of 16S rRNA within its serotypes, this method has great potential to be reapplied. Meanwhile, other genes might be needed for the bacteria whose serotypes have no difference within this gene but lie in other locations.

For antibiotic resistant detection, following work is to design a series of primers targeting to different antibiotic resistant gene and giving different melting peaks through HRM.

Last but no least, this technology will be and is trying to combine to digital PCR to achieve fastness, easiness to use, absolute quantification, high sensitivity and high specificity. Hopefully this technology will bring benefit to the modern medicine, especially to the point of care, to the remote area and developing countries where a rapid and accurate bacteria diagnosis tool can be vital to millions of lives.

CHAPTER 6 Discussion

6.1 Multiplex PCR

In qPCR, specific primer ratio of long fragment V16 to short fragment V6 gives two peaks. Changing the primer ratio will change the peak height ratio. We are able to see two clear peaks at primer ratio of 1 to 5 and 1 to 10 of long fragment primer to short fragment primer. This can be justified because short fragment binds less fluorescent and compensated by its higher primer concentration. However, sometimes the peak height of short amplicon increases in different replicate. This problem gets extremely worse in the digital PCR that sometimes we lost the long amplicon even at the primer ratio of 1:1. This makes us to think that the amplification of long fragment heavily depends on the DNA quality. When the DNA quality is low, saying DNA degradation or long DNA breaks during extraction makes the long template harder to be amplified. This problem gets worse in digital PCR because the reagent range in the reaction gets more stringent and template is much more diluted. Meanwhile in the digital PCR, as there is only one molecule in the well, the high ratio of primer for short amplicon might not be necessary and even cause the long template can't be amplified because of the primer competition.

6.2 Antibiotic resistance detection

This study only tests the antibiotic resistance gene on the genomic DNA. However, in many bacteria, the resistant gene is carried by bacteria plasmid DNA. Therefore, the above mention methods will further ask to extract both genomic and plasmid DNA for the following test. Another idea is to apply the whole cell lysis on the digital PCR which will help to retain all genetic material on the chip.

Regarding to use multiplex PCR to detect more than one antibiotic resistant gene, there are several hurdles to overcome: first whether we can design a set of primers that can give different melting peaks located at the different region from the main peak of V16, second whether multiple primers will form dimers, inhibit PCR or compete with each other to prevent the reaction and third for multi-antibiotic resistant bacteria, whether the peaks of different antibiotics will merge to the extent that are hard to distinguish.

REFERENCE

[1] Content source: Centers for Disease Control and Prevention.March 9, 2015

[2] J. Michael Janda and Sharon L. AbbottBacterial Identification for Publication: When Is Enough Enough, J Clin Microbiol. 2002 Jun; 40(6): 1887–1891.

[3] Reetta Huttunencorrespondenceemail, Jaana Syrjänen, Risto Vuento, Janne Aittoniemi. Current concepts in the diagnosis of blood stream infections. Are novel molecular methods useful in clinical practice. doi:10.1016/j.ijid.2013.04.018

[4] Yingdi Zhu, Liang Qiao, Michel Prudent. Sensitive and fast identification of bacteria in blood samples by immunoaffinity mass spectrometry for quick BSI diagnosis. Chem. Sci., 2016,7, 2987-2995

[5] Hyun Jung Chung, Cesar M. Castro, Hyungsoon Im, Hakho Lee& Ralph Weissleder. A magneto-DNA nanoparticle system for rapid detection and phenotyping of bacteria. Nature Nanotechnology 8, 369–375 (2013)

[6] Liesenfeld O, Lehmann LE, Hunfeld K-P, Kost G (2014) Molecular Diagnosis of Sepsis: New aspects and recent developments. Eur J Microbiol Immunol: in press.

[7] Mancini N, Clerici D, Diotti R, Perotti M, Ghidoli N. (2008) Molecular diagnosis of sepsis in neutropenic patients with haematological malignancies. J Med Microbiol 57: 601–604

[8] Morgan M, Marlowe E, Della-Latta P, Salimnia H, Novak-Weekley S, et al. (2010) Multicenter evaluation of a new shortened peptide nucleic acid fluorescence in situ hybridization procedure for species identification of select Gram-negative bacilli from blood cultures. J Clin Microbiol 48: 2268–2270

[9] Leggieri N, Rida A, François P, Schrenzel J (2010) Molecular diagnosis of bloodstream infections: planning to (physically) reach the bedside. Curr Opin Infect Dis 23: 311–319

[10] Tissari P, Zumla A, Tarkka E, Mero S, Savolainen L. (2010) Accurate and rapid identification of bacterial species from positive blood cultures with a DNA-based microarray platform: an observational study. Lancet 375: 224–230

[11] Samuel LP, Tibbetts RJ, Agotesku A. (2013) Evaluation of a microarray-based assay for rapid identification of Gram-positive organisms and resistance markers in positive blood cultures. J Clin Microbiol 51: 1188–1192

[12] Masayoshi Tojo, Takahiro Fujita, Yusuke Ainoda. Evaluation of an Automated Rapid Diagnostic Assay for Detection of Gram-Negative Bacteria and Their Drug-Resistance Genes in Positive Blood Cultures. PLoS One. 2014; 9(4): e94064.

[13] Fraley SI, Hardick J, Jo Masek B. Universal digital high-resolution melt: a novel approach to broad-based profiling of heterogeneous biological samples. Nucleic Acids Res. 2013 Oct;41(18):e175.

[14] Soumitesh Chakravorty, Danica Helb, Michele Burday. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods. 2007 May; 69(2): 330–339.

[15] Stephanie I. Fraley, Pornpat Athamanolap, Billie J. Masek. Nested Machine Learning Facilitates Increased Sequence Content for Large-Scale Automated High Resolution Melt Genotyping. Scientific Reports 6, Article number: 19218 (2016)

[16] Massively parallel digital high resolution melt for rapid, quantitative sequence profiling (approved).

[17] F. W. Brenner, R. G. Villar, F. J. Angulo, R. Tauxe, and B. Swaminathan. Salmonella Nomenclature. J Clin Microbiol. 2000 Jul; 38(7): 2465–2467.

[18] S. Porwollik, E. F. Boyd, C. Choy, P. Cheng, L. Florea, E. Proctor, and M. McClelland. Characterization of Salmonella enterica Subspecies I Genovars by Use of Microarrays. J Bacteriol. 2004 Sep; 186(17): 5883–5898.

[19] Gomez TM , Motarjemi Y , Miyagawa S , Käferstein FK , Stöhr K. Foodborne salmonellosis. World Health Statistics Quarterly. Rapport Trimestriel de Statistiques Sanitaires Mondiales. 1997, 50(1-2):81-89

[20] A Lupattelli, O Spigset, M J Twigg. Medication use in pregnancy: a cross-sectional, multinational web-based study. BMJ Open 2014; 4:e004365.

[21] Neal R. Chamberlain. INFECTIONS OF THE LARGE INTESTINE. 2/11/16

[22] Alexander M. Sy, Jagbir Sandhu, and Theodore Lenox. Salmonella enterica Serotype Choleraesuis Infection of the Knee and Femur in a Nonbacteremic Diabetic Patient. Case Reports in Infectious Diseases Volume 2013, Article ID 506157, 3 pages

[23] Antibiotic resistance—the need for global solutions2013

[24] Laxminarayan R1, Duse A, Wattal C, Zaidi AK. Antibiotic resistance-the need for global solutions. Lancet Infect Dis. 2013 Dec;13(12):1057-98.

[25] Jennifer Routh. NIH Funds Nine Antimicrobial Resistance Diagnostics Projects. Thursday, April 9, 2015

[26]J. T. SMITH, J. M. T. HAMILTON-MILLER. Bacterial resistance to penicillins and cephalosporins. J. Pharm. Pharmac., 1969, 21, 337-358

[27] David M. Livermore.Bacterial Resistance: Origins, Epidemiology, and Impact. Clin Infect Dis. (2003) 36 (Supplement 1): S11-S23.

[28] Harold C. Neu. The Crisis in Antibiotic Resistance. Science 21 Aug 1992:
Vol. 257, Issue 5073, pp. 1064-1073

[29] Haukur Gudnason, Martin Dufva, D.D. BangComparison. Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. Nucleic Acids Res. 2007 Oct; 35(19): e127.

[30] A.P.Magiorakos, A. Srinivasan, R. B. Carey. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clinical Microbiology and Infection Volume 18, Issue 3, pages 268–281, March 2012.