ABSTRACT

Title of Thesis: THE METHYLOME OF *YERSINIA PESTIS*

Jonathan B. Hnath, Masters of Science, 2015

Thesis Directed By: Distinguished University Professor Rita R. Colwell
Marine Estuarine Environmental Sciences
University of Maryland, College Park

During the normal cycle of transfer between its vector (the flea) and a mammalian host, *Yersinia pestis* (*Y. pestis*) is exposed to significantly different environmental conditions. Studies have shown gene expression patterns in *Y. pestis* differ significantly under these separate conditions. In many bacteria, large-scale gene expression changes are modulated by DNA methylation. To date, methylation patterns of the *Y. pestis* genome have not been examined. In this study, the methylome of *Y. pestis* was characterized and whether a change in methylation accounts for change in gene expression was determined.

The methylation pattern of the *Y. pestis* genome, having been characterized, provides a reference methylome. A comparison of the methylation state at different temperatures selected to represent vector and host conditions, showed no significant change in methylation pattern. It is concluded from this study that the methylation

pattern of the *Y. pestis* genome is not altered according to the temperature of its vector or mammalian host.

THE METHYLOME OF *YERSINIA PESTIS*

by

Jonathan B. Hnath

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Masters of Science
2015

Advisory Committee:
Professor Rita R. Colwell, Chair
Research Professor Anwar Huq
Professor Jeffrey DeStefano
Associate Professor Carlos A. Machado
Dr. Nicholas Bergman

Acknowledgements

I would like to thank my advisors Dr. Rita Colwell and Dr. Nicholas Bergman for their support and mentorship during the years of study. I would also like to thank the members of the committee for their encouraging words of wisdom throughout my time at the university.

A special thank you to Battelle National Biodefense Institute for providing tuition support and allowing me to perform the research within the laboratories. I also would like to especially thank Dr. James Burans for his support and encouraging words during my pursuit of the MS degree, Brian Ondov for his review and modifications to the R script, and the rest of the Genomics team for helping plate and review sterility plates when needed.

Most importantly, I would like to thank my wife and children for their support and love through all the time spent away from family to achieve all that has been done.

This work was funded under Contract No. HSHQDC-07-C-00020 awarded by the Department of Homeland Security (DHS) Science and Technology Directorate (S&T) for the management and operation of the National Biodefense Analysis and Countermeasures Center (NBACC), a Federally Funded Research and Development Center. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the DHS or S&T. In no event shall the DHS, NBACC, S&T, or Battelle National Biodefense Institute (BNBI) have any responsibility or liability for any use, misuse, inability to use, or reliance upon the information

contained herein. DHS does not endorse any products or commercial services mentioned in this publication.

Table of Contents

List of Tables

Table 1 - Sequencing coverage obtained for each reference sequence

Sample	Plasmid pPCP1 Coverage	Plasmid pCD1 Coverage	Plasmid pMT1 Coverage	Chromosome Coverage
Control	1080.2x	400.7x	184.08x	497.32x
28°C	3958.07x	1537.3x	802.93x	773.31x
37°C	1062.53x	453.35x	261.58x	391.07x

Table 2 - Motifs Detected in samples, the counts detected, when the QV threshold was set to 100 due to the high coverage to reduce the false positive identification of motifs

Sample	Motif	Modification Type	# Detected of Motif	# Predicted in Reference	% Detected in Sample
Control	GATC	6-mA	0	0	0%
28°C			37816	37886	99.82%
37°C			37714	37886	99.55%
Control	CCWGG	4-mC	0	0	0%
28°C			2334	5824	40.08%
37°C			0	0	0%

Table 3 -ANOVA analysis results of the each nucleotide position IPD at each comparison performed (numbers are the overall IPD percent similarity)

Sequence	28°C to 37°C	28°C to Control	37°C to Control
Plasmid pPCP1	100	99.45	99.45
Plasmid pCD1	100	99.58	99.53
Plasmid pMT1	100	99.87	99.81
chromosome	100	99.38	99.36

List of Figures

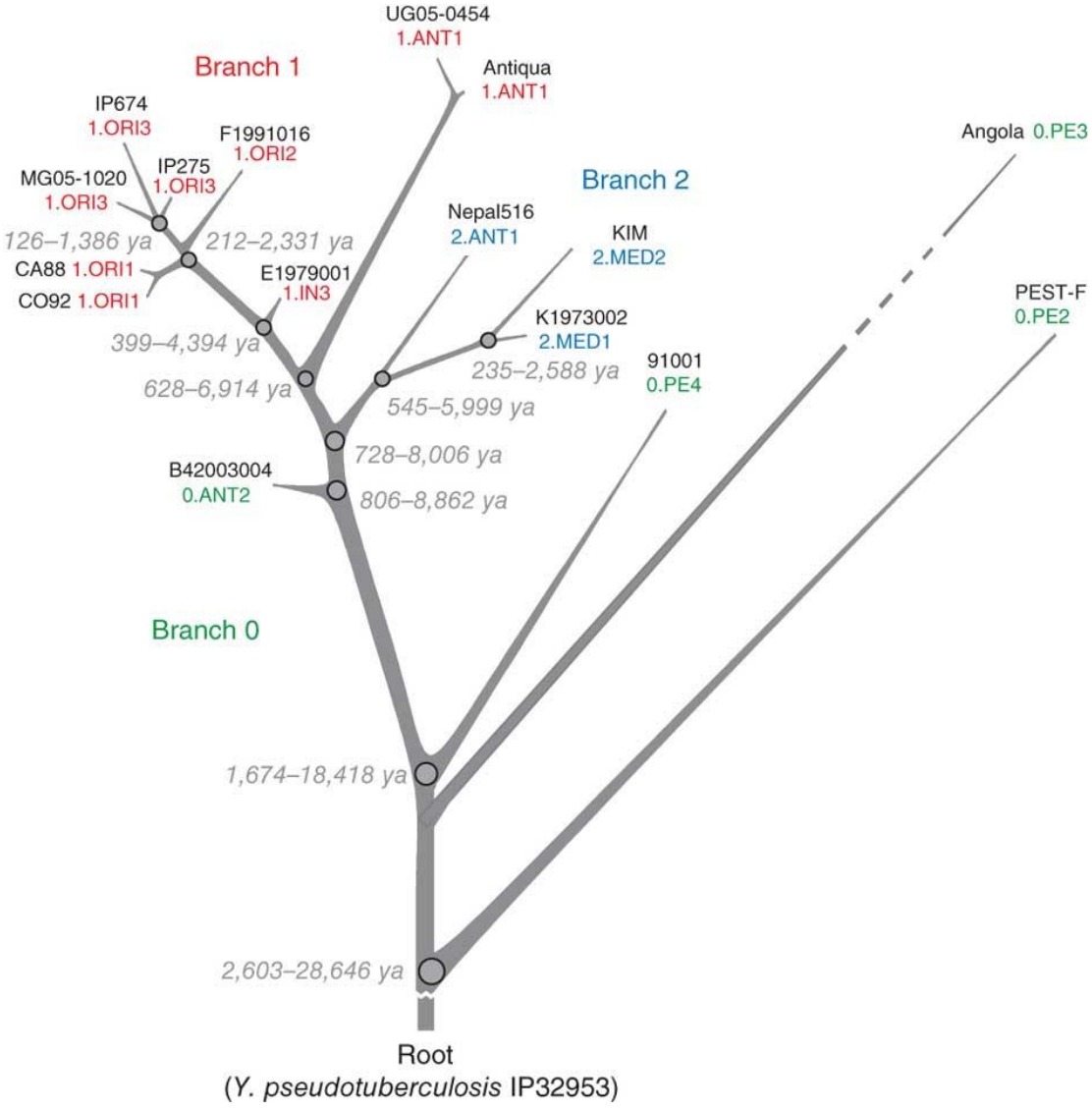

Figure 1 – Genomic maximum parsimony tree and divergence dates of *Yersinia pestis*. Black text is the name of the strain, colored text is the branch and population name, grey text is the min/max dates of divergence. Figured obtained from *Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity*[1].

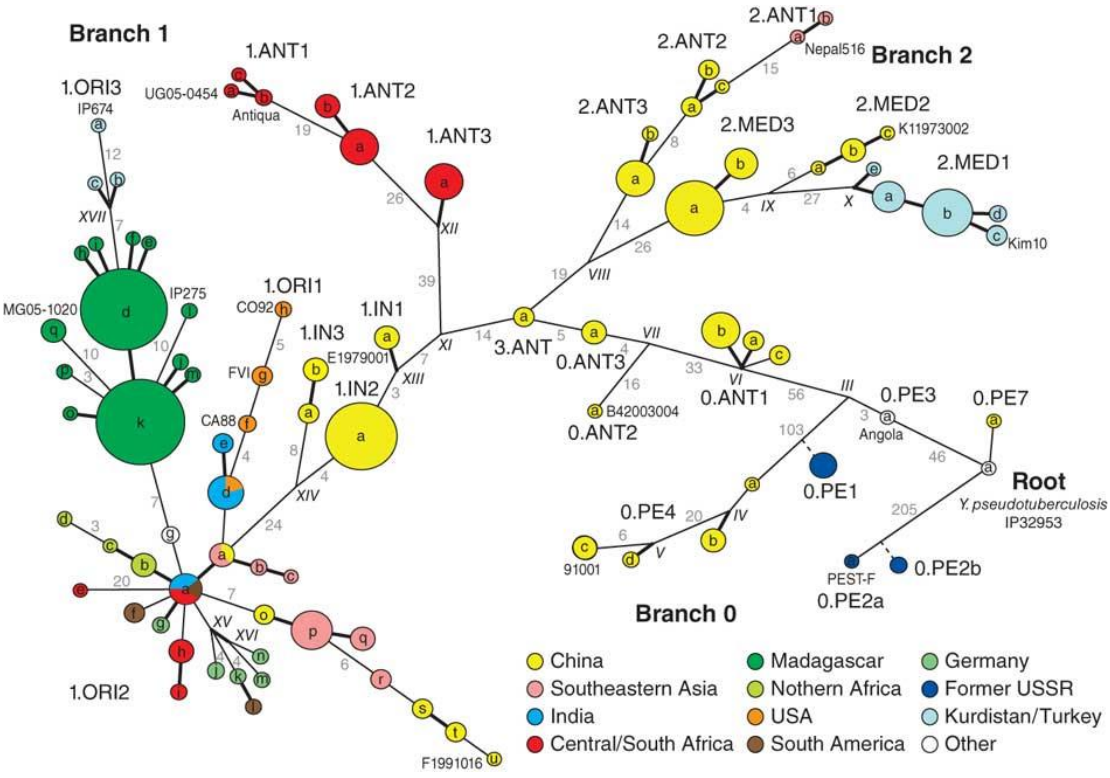

Figure 2– Fully parsimonious minimal spanning tree of the SNPs in 282 isolates of *Yersinia pestis*. Strains are listed near terminal nodes. Figured obtained from *Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity*[1].

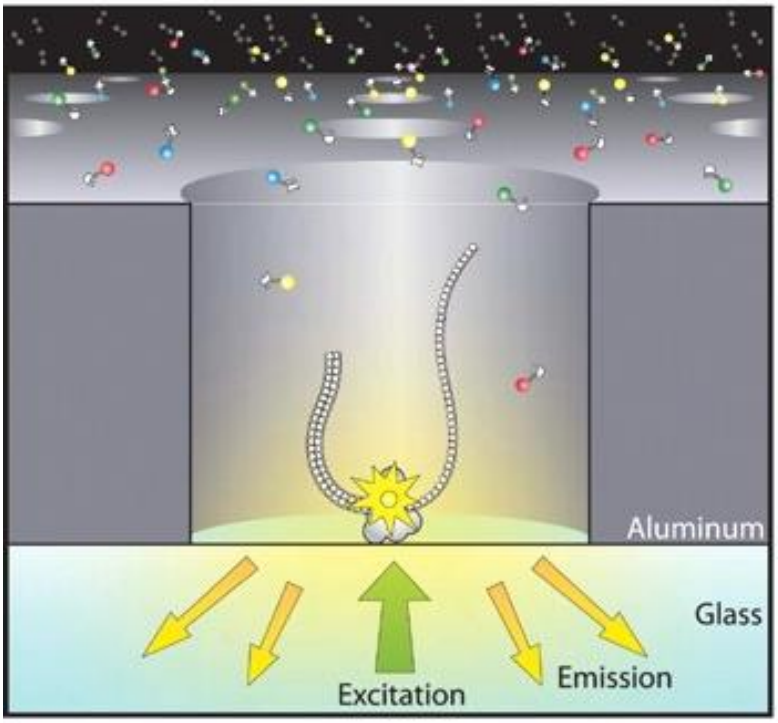

Figure 3– Image of the Pacific Biosciences zero-mode waveguide, with a molecule of DNA, bound polymerase, free dNTPs, and the method of reading the base incorporation. Figure obtained from *Real-Time DNA Sequencing from Single Polymerase Molecules*[2].

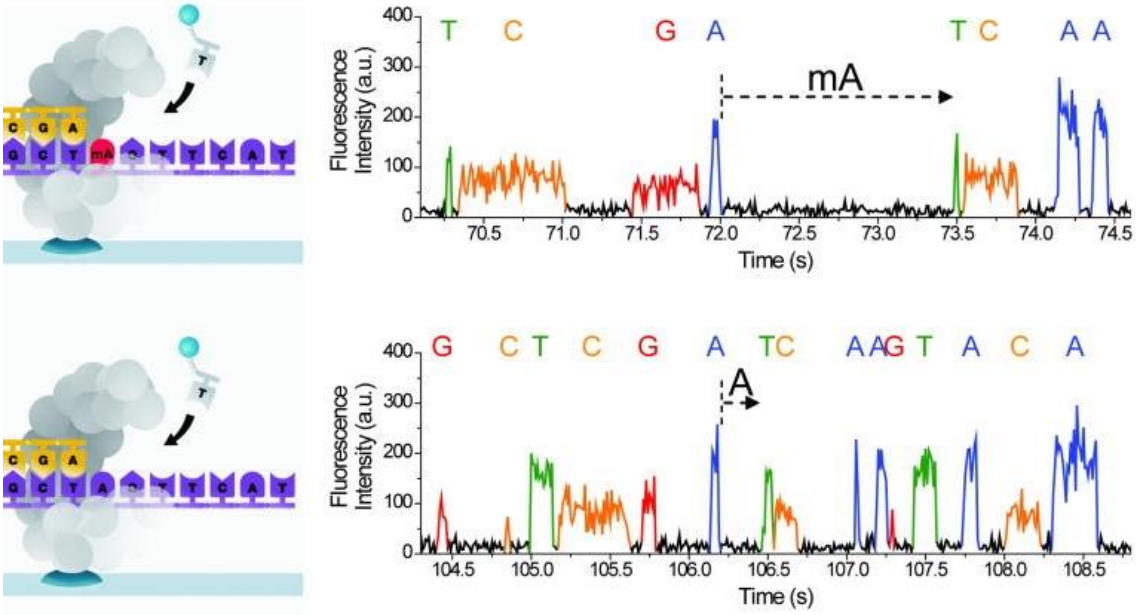

Figure 4– Image of the Pacific Biosciences detection of methylated bases. In the top portion of the image, when the base is incorporated at a methylated A location, there is an increase in the kinetics compared to the non-methylated base, lower portion. Figure obtained from *Direct detection of DNA methylation during single-molecule, real-time sequencing*[3].

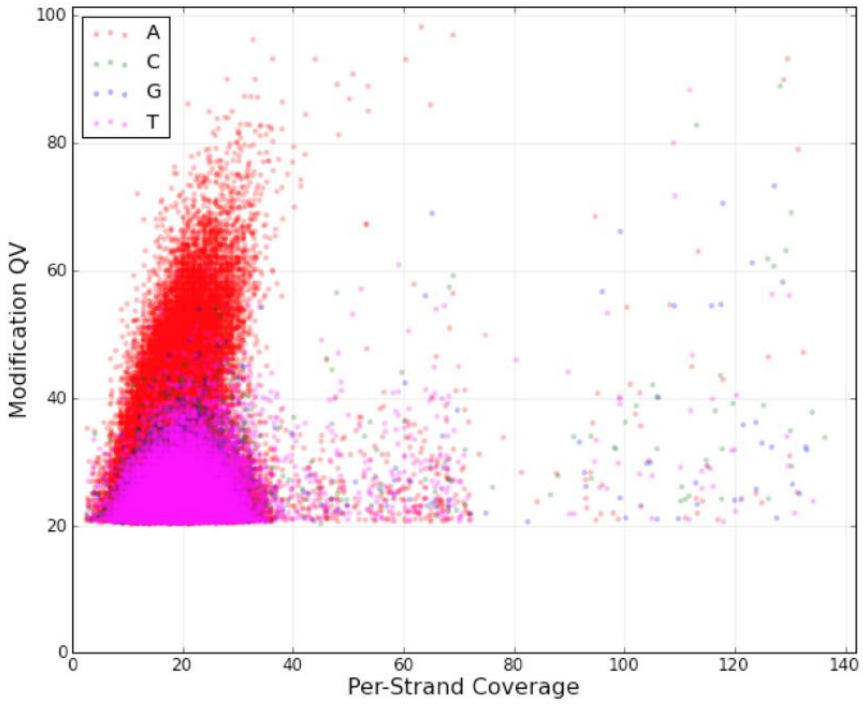

Figure 5 – Modifications quality score (QV) vs. Coverage graph for *Y. pestis* CO92 grown on TSA agar @ 30°C for 48 hrs. The peak of red scatter plots indicate an increase in methylated A bases, an indication of 6mA sites being present in the sample.

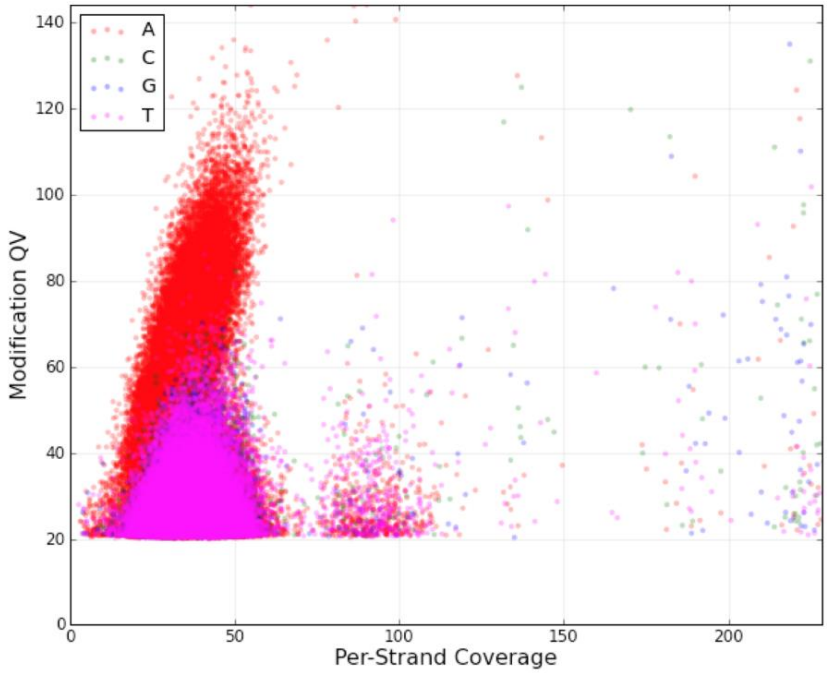

Figure 6 - Modifications quality score (QV) vs. Coverage graph for *Y. pestis* CO92 grown on BHI agar @ 35°C for 48 hrs. The peak of red scatter plots indicate an increase in methylated A bases, an indication of 6mA sites being present in the sample.

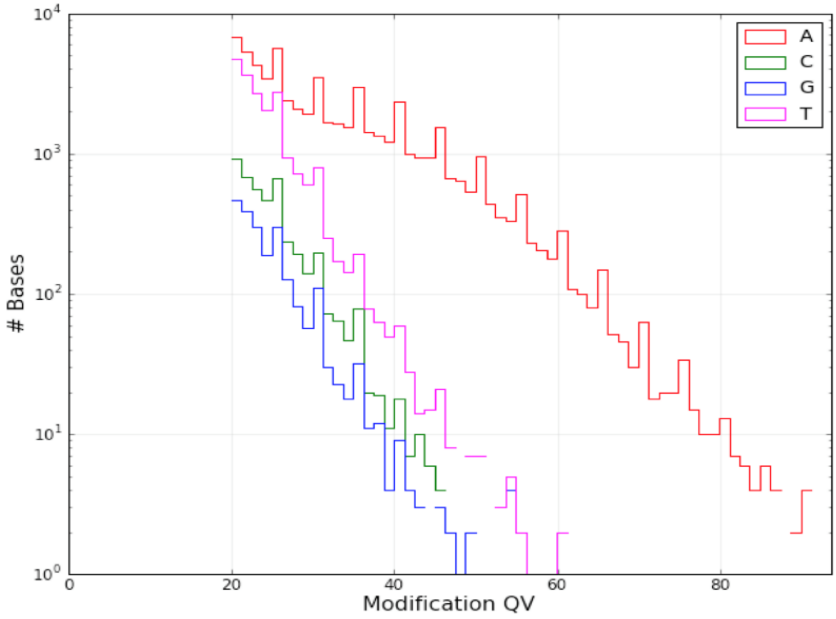

Figure 7 - Modifications quality score (QV) Histogram graph for *Y. pestis* CO92 grown on TSA agar @ 30°C for 48 hrs. The histogram shows a divergence of the A base quality scores from the other bases. This is another visualization that there are modified A bases present in the sample.

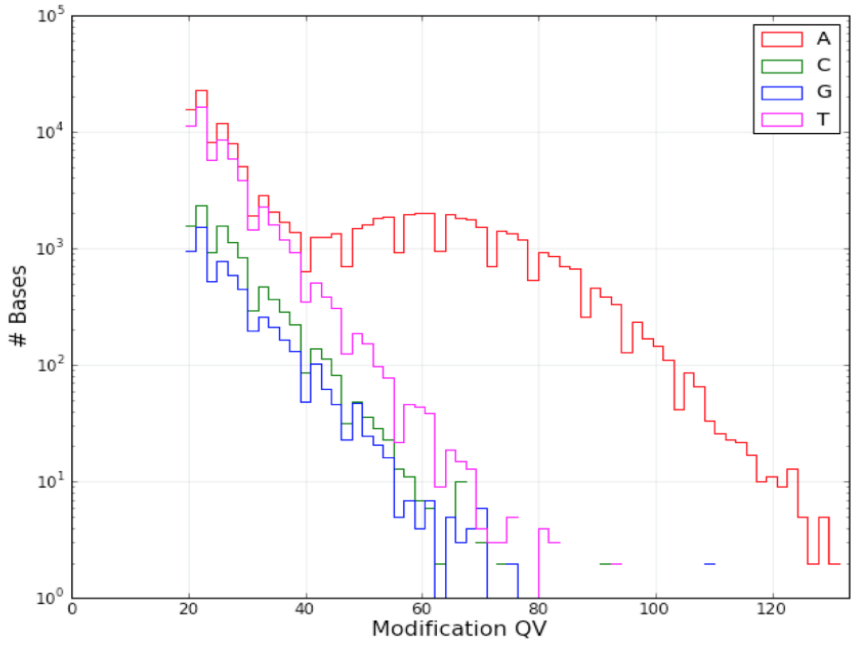

Figure 8 - Modifications quality score (QV) Histogram graph for *Y. pestis* CO92 grown on BHI agar @ 35°C for 48 hrs. The histogram shows a divergence of the A base quality scores from the other bases. This is another visualization that there are modified A bases present in the sample.

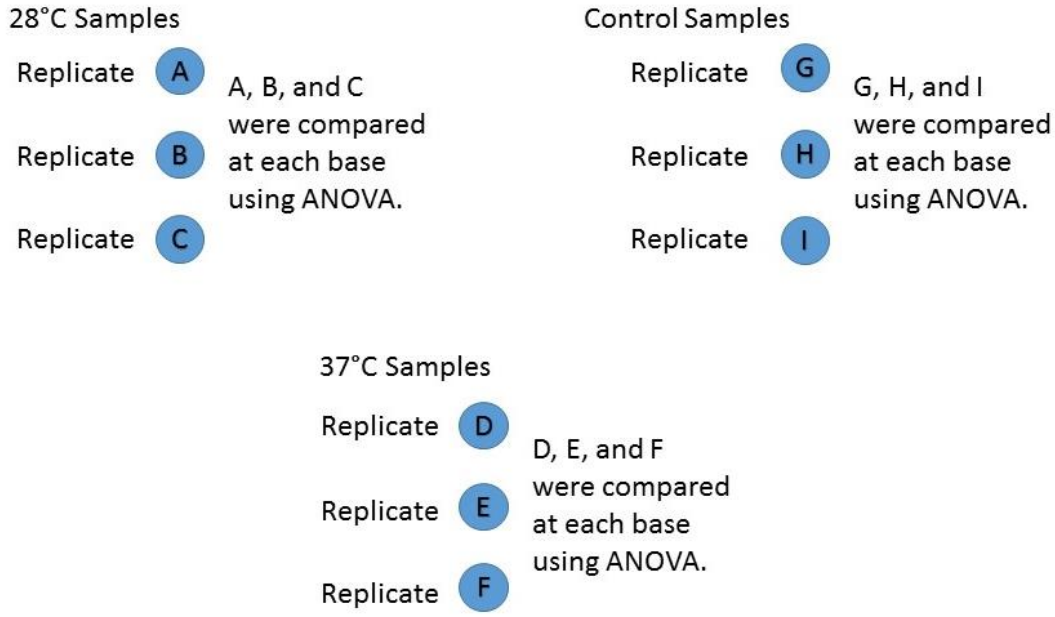

Figure 9 – Biological replicate comparison at the base level. This method was used to determine if the biological replicates were significantly similar to each other to pool together to form a larger dataset for further analysis.

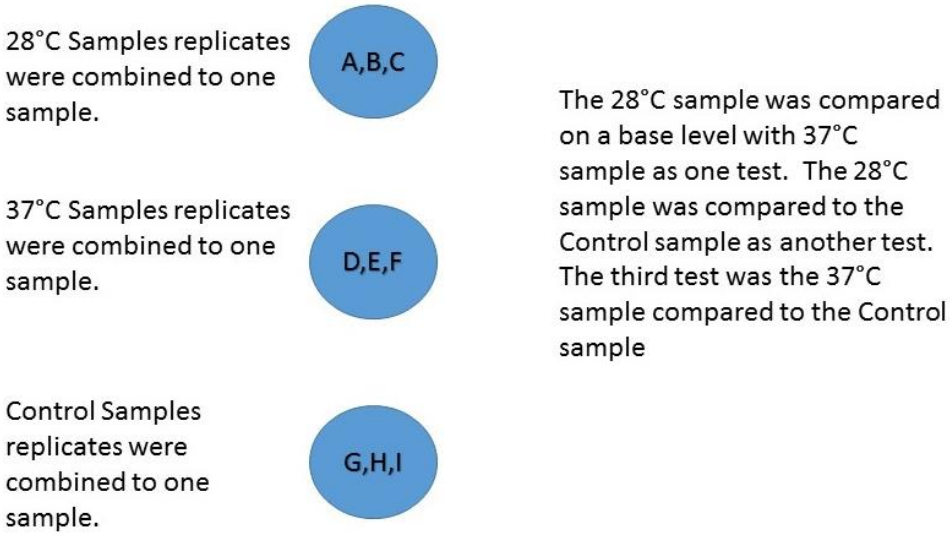

Figure 10 – Combination of biological replicates to one sample and the comparisons at the base level. This was the grouping done to create a larger pool of data for the temperature comparisons.

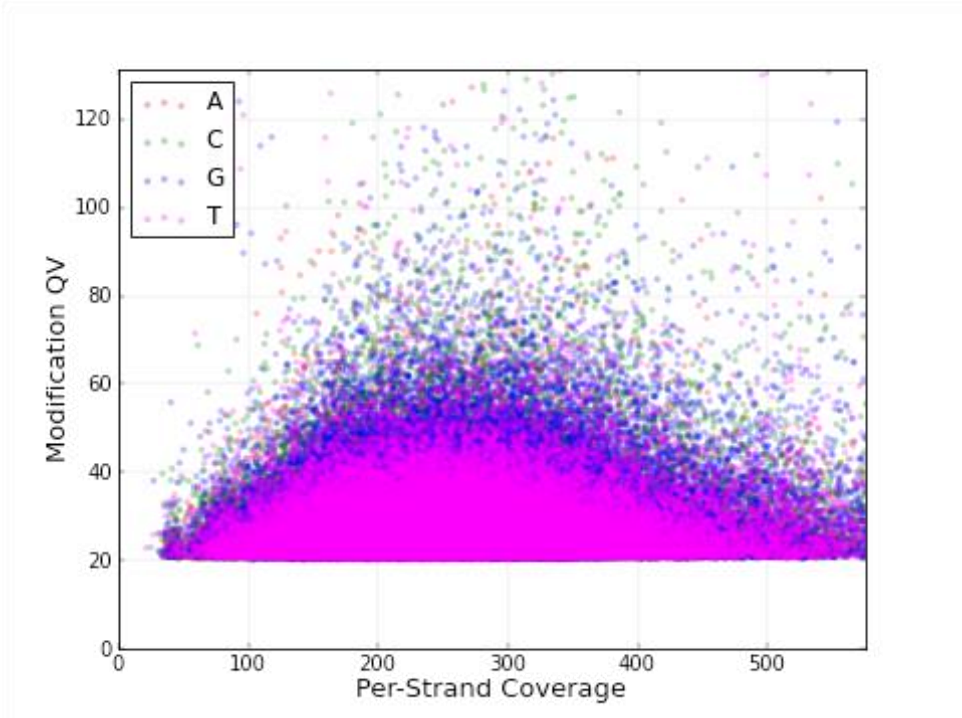

Figure 11 - Modifications quality score (QV) vs. Coverage graph for MDA treated *Y. pestis* CO92. The absence of a defined population of C or A populations indicates that there are no methylated bases present in the sample.

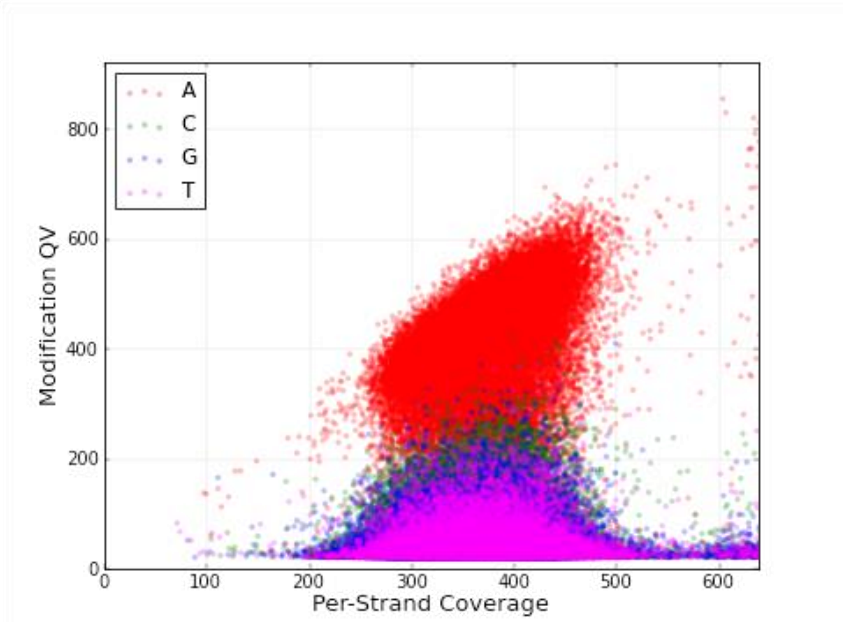

Figure 12- Modifications quality score (QV) vs. Coverage graph for *Y. pestis* CO92 grown at 28°C. The peak of red scatter plots indicate an increase in methylated A bases, an indication of 6mA sites being present in the sample. There is also a visible amount of C bases above the base but below the dense A population. This shows possible C methylation present in the sample.

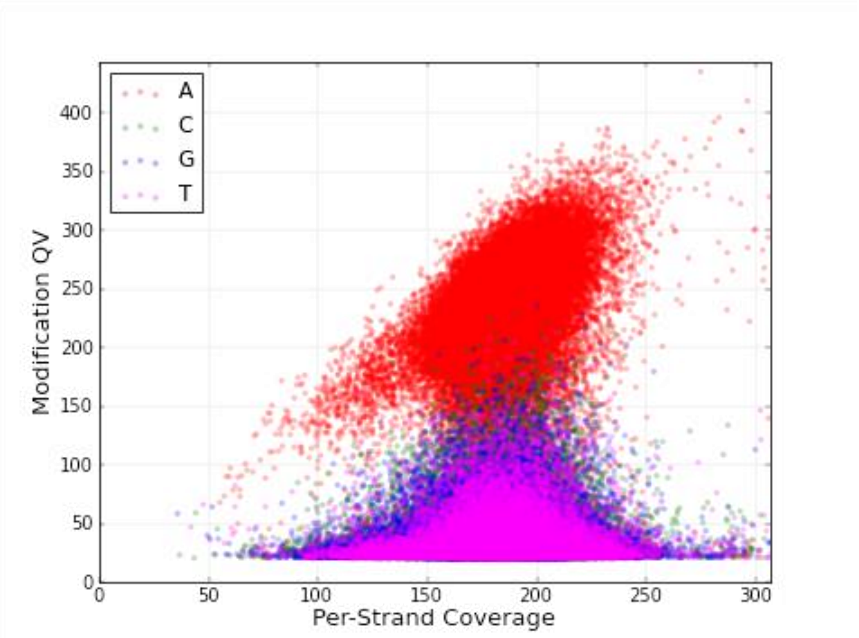

Figure 13 - Modifications quality score (QV) vs. Coverage graph for *Y. pestis* CO92 grown at 37°C. The peak of red scatter plots indicate an increase in methylated A bases, an indication of 6mA sites being present in the sample.

Introduction

Background on Yersinia pestis

Y. pestis disease

There are three clinical forms of plague: bubonic, septicemic, and pneumonic. Most cases are of the bubonic and septicemic versions of the disease. If the disease is not treated, bubonic plague mortality is approximately 40-60%, whereas septicemic and pneumonic plague are 100% fatal.

Bubonic plague received its name because this form causes swollen, tender lymph nodes, known as buboes. Buboes usually form in the lymph node closest to the site of the bite from a flea, from which the infection is transmitted. Other symptoms include fever, headache, chills, and weakness, which usually last between two to eight days[4-6].

Septicemic plague is defined when blood culture is positive and the symptoms are similar to those of gram-negative bacterial septicemias. Septicemic plague has a 30-50% mortality rate, even with antibiotic treatment. The incubation period for septicemic plague is one to four days[5-6].

Pneumonic plague is very rare, but extremely deadly. It is estimated that the infectious dose is as low as 100 to 500 organisms. The methods of transmission are inhalation of the bacteria via respiratory droplets of infected individuals or animals. Accidental inhalation of plague bacteria in the laboratory has also been reported[5-6].

History of *Y. pestis* outbreaks

Three pandemics have been described in the history of *Y. pestis*, the first being the Justinian plague, 541-767 AD. During this pandemic that occurred in the Eygptian and the Mediterranean basins, approximately 100 million people died from the disease. The second and most famous pandemic is the Black Plague that occurred during the Middle Ages (from 1346 to the 1800s) and is believed to have killed approximately a third of the population of Europe[5-7]. The third pandemic began in the mid-1800s[5-7]. With the exception of Australia, *Y. pestis* is globally endemic and today the disease results from contact with infected wild animal reservoirs, rather than being transmitted by infected rats. Thus, human cases of plague usually are preceded by an epizootic outbreak in the geographic region of an epidemic[5-6].

Yersinia pestis, the causative agent

Y. pestis, the causative agent of plague, is a Gram-negative bacterial pathogen discovered by Alexandre Yersin during his studies of the fluid from enlarged lymph nodes of plague victims. At the time Yersin was able to prove only that the disease was caused by a gram-negative bacterium and it was Paul-Louis Simond who made the discovery that *Y. pestis* was transmitted by rat fleas[5-6].

Y. pestis has been shown to survive in soil for long periods of time and in his studies Yersin was able to isolate *Y. pestis* from the soil around homes of infected individuals[8]. In 2008, a study was done to examine the ability of *Y. pestis* to survive in soil, showing *Y. pestis* could remain viable in soil for up to 40 weeks[8]. However, *Y.pestis* will die rapidly in soil at temperatures of 40°C or greater and also in desiccated soil[6,8].

Modes of transmission

Rodents are the primary reservoir of *Y. pestis* in the wild and fleas are the principal vector of *Y. pestis*. While only a few species of flea are known to serve as a vector for the disease, over 100 species have been shown to be capable of harboring the bacterium[9]. Typically, *Y. pestis* is transmitted between hosts when a flea acquires the bacteria while feeding upon an infected rodent and moves on to feed on an uninfected rodent. The disease often spreads outside its typical cycle when an infected flea feeds on another mammalian host, including humans. Other modes of infection include aerosols and handling contaminated fluids or tissue.

The most commonly considered transmission system is the proventricular blockage model. In this model, fleas become infected by feeding on the blood of infected rodents. After feeding, *Y. pestis* begins to grow and multiply within the gut of the flea. Once the bacteria have multiplied sufficiently to block the gut of the flea, the flea will regurgitate its meal, causing the bacteria to be introduced into the bite wound during feeding[9-12]. Even with this being the most widely accepted model, a study carried out in 2006 presented data demonstrating transmission of *Y. pestis* was efficient, even when flea guts were not blocked and suggesting there are other mechanisms of early-phase transmissions[13].

A mechanistic transmission model has been proposed that is similar to sharing an infected needle. When fleas feed on infected hosts, blood left on the mouth part of the flea is infected with *Y. pestis*. When the flea takes another blood meal, the new host is infected with the contaminated blood present on the mouth of the flea[12].

Temperature is also believed to have an impact on the ability of *Y. pestis* to spread and be transmitted. A study reported in 2011 showed that *Y. pestis* can be transmitted even at high temperatures. It was also concluded that plague outbreaks would be hard to maintain when temperatures are in the range of 27°C -30°C. Thus, it is concluded that other factors contribute to the success of transmission at high temperatures[14].

Formation of biofilms has been linked to the blockage model. The ability of *Y. pestis* to produce biofilms was discovered when the bacterium was grown on Congo red agar at 28°C. The colonies are pigmented, but when grown at 37°C pigmentation is blocked, a characteristic phenomenom of bacterial biofilms. In the case of *Y. pestis*, biofilm formation has been shown to play a vital role in transmission from the flea. The bacterium, when exposed to the lower temperature of the flea, will form a biofilm matrix which, in turn, allows the bacterium to colonize the proventricular spines of the flea, thereby aiding transmission[15].

Genetic information

The genome of *Y. pestis* comprises a 4.6 MB chromosome and three plasmids -- pCD1 (~70kb), pPCP1 (~9.6 kb), and pMT1 (~100 kb)[16]. The known virulence genes reside predominantly on the plasmids[17]. *Y. pestis* has been grouped into four biotypes: Antiqua, Medievalis, Orientalis, and Microtus. Assignment to biotype is determined by ability of the isolates to reduce nitrate and ferment sugars[1,18]. Sequence data indicate *Y. pestis* has undergone many structural changes in its genome that allow for adaptation, including (but not limited) to horizontal gene transfer, gene loss, recombination, and mutation[9,16].

Plasmid pCD1 is also known as pLCR, pYV, or the low calcium response plasmid, and contains virulence genes of the type III secretion[17-18]. Proteins for which the genes code are referred to as *Yersinia* outer proteins or Yops for short. There are a few proteins that interfere with phagocytosis and these include GTPase activating protein (YopE), and tyrosine phosphatase (YopH), both of which are antiphagocytic, and a serine threonine kinase (YopO/YpkA). YopM, located on this plasmid, is linked to a transcription event of cytokine genes[19]. The YopJ/P gene is also located on this plasmid and has been associated with inhibition of proinflammatory cytokines, tumor necrosis factor-α, and production and induction of macrophage apoptosis[17]. A cytotoxin (YopT) has been shown to cause filament disruption.

There are six additional genes and these are involved in control and translocation of Yops and include yopN, yopB, yopD, tyeA, lcrG, and lcrV[17]. The LCR genes on the *Y. pestis* pCD1 plasmid have been shown to be 98% identical to the two *Yersinia enterocolitica* plasmids[17]. It has also been determined that the YopM gene contains two extra copies of a repeat sequence in the *Y. pestis* pCD1 plasmid, compared to the two *Yersinia enterocolitica* plasmids[17].

Plasmid pPCP1 contains the plasminogen activator (Pla or Pst) which has been shown to be essential for *Y. pestis* virulence when transmitted by flea bite[17]. Plasmid pMT1, also known as pFra, encodes the gene for the murine toxin and the fraction 1 (F1) capsular antigen. The murine toxin is a toxin required for survival of *Y. pestis* in the flea and the F1 capsular antigen is thought to be required for full

virulence of *Y. pestis*[17]. Both plasmids, pPCP1 and pMT1, are specific to *Y. pestis* and are not found or shared by any other enteropathogenic Yersiniae[17].

Y. pestis genome analysis suggests it evolved from *Yersinia pseudotuberculosis* as early as 1,500-20,000 years ago[1,16]. See Figure 1. About 13% of the *Yersinia pseudotuberculosis* genome is not present in the genome of *Y. pestis*[18]. *Y. pestis* is believed to have spread through the trade routes that originate in China. During traveling across and populating into other regions, *Y. pestis* evolved into different lineages[18]. A review of all *Y. pestis* isolates that have been sequenced and their unique SNPs, presented in the tree, shown in Figure 2, suggests that *Y. pestis* evolved in China and spread to other locations, supporting the trade route theory[1].

DNA methylation

Known Methylation roles

DNA methylation has been identified as being important for many bacterial processes, including mismatch repair, regulation of gene expression, and pathogenicity[20-22]. Genome methylation in bacteria usually occurs as 5-methylcytosine (5-mC), N^4-methylcytosine (4-mC), or N^6-methyladenine (6-mA). The most common form of methylation in bacteria is 6-mA of the sequence GATC and this process is catalyzed by DNA adenine methyltransferase (*dam*)[20]. For the above mentioned processes, the events are signaled by the hemimethylated state of DNA[21]. Hemimethylated DNA is where one strand of DNA is methylated while the other is not. Dam moves along the DNA in a linear fashion, methylating 5'-GATC-3' sites at a rate of 20-100 sites per minute[20-21].

Methylation impact on *Yersinia pestis*

A mutation to the *dam* gene has been shown to affect virulence of *Y. pestis*[22]. In one study, a version of *Y. pestis* was constructed in which the *dam* gene was inactivated. Mice were individually exposed to the mutant or the wild type strain. Those mice exposed to the mutant survived, indicating the strain had been attenuated and subsequently, when the mice were exposed to the wild type strain, they also survived, showing the mutant induced protection as well. From these results, it was concluded that the *dam* gene plays a role in the virulence of *Y. pestis*. This experiment was also carried out with *Y. pseudotuberculosis* in mice, showing protection against exposure to *Y. pestis* after the mice had been dosed with the *dam* mutant strain of *Y. pseudotuberculosis*[23].

Overproduction of the *dam* gene product was shown to attenuate strains of *Yersinia*[24-25]. This was tested by exposing mice to the strains that were overproducing product of the *dam* gene. The mice survived exposure, hence the conclusion that the strain had been attenuated. Mice subsequently exposed to the virulent strain survived, suggesting protection. It should be noted that overproduction has an impact on secretion of the Yop genes, which are known to be essential virulence factors[24-25].

Sequencing Technology

Pacific BioSciences Technology

The SMRT® technology of Pacific Biosciences (PacBio) utilizes DNA polymerase along with phospholinked dNTPs to capture DNA sequences in real-time[2]. The technology utilizes wells in a chip called zero-mode waveguide (ZMW) to

capture a single molecule of DNA. While the bases are incorporated there is a pulse which is the elevation of fluorescence output from the ZMW[2]. With this technology, strands of DNA can be sequenced in real-time and these strands of DNA can be long reads (a limitation in other Next-Generation sequencing machines). A ZMW depiction is shown in Figure 3. Within this image, there is a single strand of DNA located in the ZMW. While the polymerase is incorporating a dNTP, emission of the phospholinked dNTP creates a pulse. Based on color and pulse duration, the sequence can be determined[2].

Methylation Detection

PacBio technology uses kinetic measurements to identify positions in a DNA sequence that may contain base modifications, such as methylation, because modification affects the rate of nucleotide addition by the polymerase during the sequencing reaction. The interpulse distance (IPD) is a measurement representing the time between the pulse corresponding to the addition of nucleotide_n in a growing strand of DNA and the pulse of nucleotide_{n+1}. Numerous studies have shown that an IPD is significantly longer if nucleotide_{n+1} is opposite a methylated template base[3]. Figure 4 shows a representation of how the kinetics are affected when a base is incorporated in the presence of methylation versus a non-methylated base.

Objectives of this Study

The overall goal of this project was to improve understanding of how DNA methylation impacts the *Yersinia pestis* genome. Little is known about how DNA methylation affects virulence and studies of DNA methylation in *Y. pestis* are limited.

Therefore, the research proposed here was to analyze DNA methylation in *Y. pestis* and determine the potential effect, if any, on pathogenicity. *Y. pestis* DNA methylation patterns were analyzed after growth at the two classic temperatures and media mimicing the flea (vector) and rodent/human (host) growth conditions. The objective, thus, was to determine whether *Y. pestis* has a small (localized) or wide (whole genome) scale pattern of DNA methylation within its genome.

The overall aim was to determine if the *Y. pestis* genome contains methylated bases detectable within its genome and to define genome-wide methylation patterns occuring under relevant growth conditions. It is not known what percentage of the *Y. pestis* genome is methylated. Methylation is an important part of several systems and processes, such as defense mechanisms, gene expression, cell cycle and DNA replication, DNA damage and repair, and pathogenicity. To examine methylation patterns, different growth conditions are employed and methylomes compared to determine significant differences, if any, between the DNA methylation patterns.

In this study, it was hypothesized that the Y. pestis genome is methylated, based on results obtained in an earlier study using the dam mutant strain, and a significant difference in methylation patterns for *Y. pestis* under different growth conditions would be found.

Methods and Materials

Bacterial Strain Employed in Study

Y. pestis CO92, containing the three intact plasmids pCD1, pMT1, and pPCP1, was selected for analysis because it is a virulent strain[16].

Genome Sequencing

Y. pestis CO92 was grown on 5% Sheep Blood Agar plates at 28°C for 2-3 days to revive the culture from a glycerol stock. A single colony was selected and transferred to 1 mL of Trypticase Soy Broth (TSB). A total of nine Congo Red Agar plates were inoculated with 100 µL of the TSB culture. Three of the Congo Red Agar plates were incubated at 37°C and the remaining six plates were incubated at 28°C for four days. Temperatures of 28°C and 37°C, are temperatures routinely used to culture *Y. pestis* as they represent vector and host temperatures, respectively. DNA was extracted from each plate using Wizard® Genomic DNA Purification Kit (Promega). Of the nine DNA preparations, three were subjected to Multiple Displacement Amplification using the illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kits (GE Healthcare Life Sciences) to remove all methylation from the DNA.

Pacific Biosciences (PacBio) libraries were prepared following Procedure & Checklist 10 kb Template Preparation and Sequencing (with Low-Input DNA) (http://www.smrtcommunity.com/SampleNet/Sample-Prep) using the SMRTbell™

Template Prep Kit 1.0 (Pacific Biosciences). The libraries were sequenced on the PacBio RSII instrument using DNA/Polymerase Binding Kit P4 (Pacific Biosciences), DNA Sequencing Reagents 2.0 (Pacific Biosciences), and SMRT® Cell V3 (Pacific Biosciences).

Detection of DNA Methylation

Detection of DNA methylation was carried out using the PacBio SMRT Analysis pipeline (http://www.pacb.com/devnet/). Each modified base position was determined using the RS_Modification_and_Motif_Analysis protocol within the PacBio SMRT® Portal SMRT® Analysis v2.2.0. The minimum modification QV score was adjusted to 100 from the default 30 because of the high coverage obtained.

DNA Methylation Comparison

The cmp.h5 file that resulted from alignment to the reference sequences (*Y. pestis* CO92 complete chromosome, NC_003143.1; *Y. pestis* CO92 complete genome pCD1, NC_003131.1; *Y. pestis* CO92 complete genome pMT1, NC_003134.1, *Y. pestis* CO92 complete genome pPCP1, NC_003132.1) was processed through the PacBio python script cmph5tools.py to extract all reads associated with each reference (command line was: source /dir/smrtanalysis/current/etc/setup.sh cmph5tools.py select --groupBy Reference aligned_reads.cmp.h5). Each resulting cmp.h5 file was processed with an in-house bash script that utilized multiple cores to process the large reference sequences efficiently. The bash script processed an in-house R script which utilized an available PacBio R script source to parse through each base of the reference sequence, obtain the IPD's for each read at that base, and

then perform an ANOVA test. ANOVA was used to compare sample group tests as shown in Figures 9 and 10.

Results

Preliminary Analysis

Y. pestis grown at both temperatures (approximating temperatures of the flea vector and the human host) had a genome containing extensive DNA methylation. The IPD distributions in the *Y. pestis* sequence data were examined using PacBio software, revealing a population of modified adenines in the genomic DNA samples from both sets of DNA, i.e., from cells grown at 30°C on Trypticase Soy agar (TSA) and 35°C on Brain Heart Infusion (BHI) agar. Figures 5 and 6 show quality values assigned to base modifications by the Pacific Biosciences software; methylation at the adenine bases is indicated by high Modification QV scores for the subpopulation of red "A" residues. Modifications were observed at both 30°C on Trypticase Soy agar (TSA) (Figure 5) and those grown at 35°C on Brain Heart Infusion (BHI) agar (Figure 6). Figures 7 and 8 show the associated quality score histogram plots and the skew of adenines toward higher QV scores than other bases, namely the presense of extensive methylation of the adenine bases. Although qualitative comparison could be made, the sequence coverage was insufficient for quantitative comparison of patterns for the samples. Therefore, the focus was on obtaining additional data on *Y. pestis* grown at the two different temperatures and performing quantitative comparative analysis.

Results

Methylome Identification

The methylome of *Y. pestis* CO92 was determined by analyzing Pacific Biosciences sequencing data and identifying methylation motifs. The data from the 28°C sample showed two different motifs with large coverage of the sequence. The motifs were 5'-G**A**TC-3' and 5'-C**C**WTGG-3', with base modifications of 6-mA and 4-mC, respectively. Data for the 37°C sample revealed a single motif most likely due to lower coverage, with the motif of 5'-G**A**TC-3' and a base modification of m6A. The control sample yielded no motifs, which was expected since MDA removed all methylation from the nucleotides prior to preparing the PacBio libraries. Table 1 shows depth of coverage obtained and Table 2 a breakdown of the motifs, number of detections of motif, predicted number of motifs in the reference sequence, and overall percent of the predicted motifs detected.

Methylome Comparison

Y. pestis CO92 control and experimental replicates were sequenced using the Pacific Biosciences platform. Average depth was greater than 200x coverage of the chromosome. Table 1 shows the complete listing of coverage for the chromosome and three plasmids. The IPDs at each base (for all positions in the genome) for each biological replicate were tested for significant differences using ANOVA, with a *p*-value threshold of $p < 1E^{-8}$, as shown in Figure 9. This result is akin to a quality control step, where the degree of relatedness for the replicates within a set is determined. The biological replicates in each of the groups (Control, 28°C, and

37°C) were similar (within each set) at the base level, indicating methylation patterns of the replicates were not significantly different. Based on this finding, the data within replicates were merged to allow a larger comparison and avoid the possibility that a single replicate would skew the analysis. Comparison of all data collected for the sequences obtained from cultures grown at the two temperatures is shown in Figure 10. Results of the analysis showed that, although comparison of unmethylated samples from each of the growth temperatures identified a large number of methylated adenosines, the methylation profiles from the DNA of cells grown at the two different temperatures were indistinguishable—there were no positions in the genome at which one profile was significantly different. Even though the motif finder was able to identify a second motif in the 28°C sample and not the 37°C sample, IPD values for the two samples were not significantly different, taken as proof that the coverage allowed the motif finder to identify a different motif with confidence. As shown in Figures 11-13, modification profiles for the samples showed an obvious difference between control and test samples. However, the differences between the 28°C and 37°C profiles, though visually different, were not statistically significant. The percentage of similar IPD values, when subjected to ANOVA analysis, showed the 28°C and 37°C DNA sequences were 100% identical at the base level for both the chromosome and plasmids. When compared to the control, both the 28°C and 37°C sequences were >99% identical to the IPD values of the unmethylated control. Table 3 provides a complete breakdown of the percent similarity between tests performed for the results obtained for two growth temperatures.

Discussion

Yersinia pestis methylome

The methylome of *Y. pestis* was successfully determined and consists of 5'-G**A**TC-3' and 5'-C**C**WTGG-3' motifs, with base modification of 6-mA and 4-mC, respectively. As shown in Table 3, the second motif was not detected in the 37°C sample, due to different coverage. If the QV threshold had been adjusted to be less stringent for the 37°C sample, the motif would have been identified. However, with high coverage, the QV threshold should have been set more stringently to eliminate a false motif from being identified as a potential motif. Preliminary results were obtained using the Tet1 method to determine if any 5-mC motifs were present and the results showed none.

Methylome pattern comparison

A comparison of the methylations obtained for *Y. pestis* grown at two temperatures, representing temperatures of the vector and host, showed no significant difference in their methylation patterns. The conclusion is that *Y. pestis* does not alter its methylation pattern to control growth at different temperatures and, more likely, utilizes methylation for other purposes. One such purpose may be regulation of gene expression. To test this hypothesis, an RNA-seq experiment, with analysis at the sequence level, should show changes in gene expression patterns for *Y. pestis* grown at different temperatures. This, in fact was attempted previously, but was done using microarrays which have limitations[22]. An RNA-seq experiment is preferred since all

of the RNA present would be analyzed, compared with the limited known sequences on a microarray. A *dam* mutant can be utilized to observe the impact on gene regulation at the two temperatures. Lastly, overproduction of the *dam* gene could be utilized to observe effects, if any, on gene expression from overproduction of *dam.* This has been done with *Y. pseudotuberculosis* and has shown that virulence factors are inhibited[24-25]. Obtaining better understanding of *dam* gene regulation of genes in *Y. pestis* should provide valuable insight with respect to finding better antibiotics, antimicrobial cleaners, and potential vaccines against plague, a disease which remains a global threat.

In the course of this study, a multiprocessor software tool that runs comparison analysis and reports the percentage of bases with similar IPD profiles, was developed. The tool significantly decreases hands-on time and is available upon request.

Bibliography

1. *Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity*. **Morelli, Giovanna, et al.** 42, October 2010, Nature Genetics, pp. 1140-1143.

2. *Real-Time DNA Sequencing from Single Polymerase Molecules.* **Eid, John, et al.** 2009, Science, pp. 133-138.

3. *Direct detection of DNA methylation during single-molecule, real-time sequencing.* **Flusberg, Benjamin A., et al.** June 2010, Nature Methods, pp. 461-465.

4. **U.S. Department of Health and Human Services.** *Biosafety in Microbiological and Biomedical Laboratories.* 5th. 2009. pp. 159-160.

5. *Yersina pestis and the Plague.* **Rollins, MD, Sarah E, Rollins, PhD, Sean M and Ryan, MD, Edward T.** 2003, American Society for Clinical Pathology, Vol. 119, pp. S78-S85.

6. *Yersina pestis - Etiologic Agent of Plague.* **Perry, Robert D. and Fetherston, Jacqueline D.** Jan 1997, Clinical Microbiology Reviews, Vol. 10, pp. 35-66.

7. **Plague. *CDC.* [Online] http://www.cdc.gov/plague/.**

8. *Long-term persistence of virulent Yersinia pestis in soil.* **Ayyadurai, Saravanan, et al.** 2008, Microbiology, Vol. 154, pp. 2865-2871.

9. *Comparative and evolutionary genomics of Yersinia pestis.* **Zhou, Dongsheng, et al.** 6, 2004, Microbes and Infection, pp. 1226-1234.

10. *Quantitative Competitive PCR as a Technique for Exploring Flea-Yersinia pestis Dynamics.* **Engelthaler, David M., et al.** 62(5), 2000, The American Society of Tropical Medicine and Hygiene, pp. 552-560.

11. *Observations on the mechanism of the transmission of plague by fleas.* **Bacot, A. W. and Martin, C. J.** 1914, Journal of Hygiene, pp. 423-439.

12. *The Evolution of Flea-borne Transmission in Yersinia pestis.* **Hinnebusch, B. Joseph.** 7(2), Jul 2005, Current Issues in Molecular Biology, pp. 197-212.

13. *Early-phase transmission of Yersinia pestis by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics.* **Eisen, Rebecca J., et al.** 2006, PNAS, pp. 15380-15385.

14. *Effects of temperature on the transmission of Yersinia Pestis by the flea, Xenopsylla Cheopis, in the late phase period.* **Schotthoefer, Anna M., et al.** 4:191, 2011, Parsites & Vectors.

15. *Yersinia pestis Biofilm in the Flea Vector and Its Role in the Transmission of Plague.* **Hinnebusch, B. J. and Erickson, D. L.** 2008, Current Topics in Microbiology and Immunology, pp. 229-248.

16. *Genome sequence of Yersinia pestis, the causative agent of plague.* **Parkhill, J., et al.** October 4, 2001, Nature, Vol. 413, pp. 523-527.

17. *Current Trends in Plague Research: From Genomics to Virulence.* **Huang, Xiao-Zhe, Nikolich, Mikeljon P. and Lindler, Luther E.** 3, 2006, Clinical Medicine and Research, Vol. 4, pp. 189-199.

18. *Plague in the genomic area.* **Drancourt, M.** 18, 2012, Clinical Microbiology and Infection, pp. 224-230.

19. *Toward a molecular pathogenic pathyway for Yersinia pestis YopM.* **Uittenbogaard, Annette M., et al.** December 2012, Frontiers in Cellular and Infection Microbiology, Vol. 2:155.

20. *Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more.* **Marinus, Martin G. and Casadesus, Josep.** 33(3), May 2009, FEMS Microbiology Review, pp. 488-503.

21. *Epigenetic Gene Regulation in the Bacterial World.* **Casadesus, Josep and Low, David.** 3, 2006, Microbiology and Molecular Biology Reviews, Vol. 70, pp. 830-856.

22. *A dam mutant of Yersinia pestis is attenuated and induces protection against plague.* **Robinson, Victoria L., Oyston, Petra C.F. and Titball, Richard W.** 2005, FEMS Microbiology Letters, pp. 251-256.

23. *Oral immunization with a dam mutant of Yersinia pseudotuberculosis protects against plague.* **Taylor, Victoria L., Titball, Richard W. and Oyston, Petra C. F.** 2005, Microbiology, pp. 1919-1926.

24. *DNA Adenine Methylase Is Essential for Viability and Plays a Role in the Pathogenesis of Yersinia pseudotuberculosis and Vibrio cholerae.* **Julio, Steven M., et al.** Dec 2001, Infection and Immunity, pp. 7610-7615.

25. *DNA Adenine Methylase Overproduction in Yersinia pseudotuberculosis Alters YopE Expression and Secretion and Host Immune Responses to Infection.* **Julio, Steven M., et al.** Feb 2002, Infection and Immunity, pp. 1006-1009.