2015

Gap Closure to Complete the Genome Assembly for Staphylococcus agnetis

Joseph A. Koon II
Ouachita Baptist University

SENIOR THESIS APPROVAL

This Honors thesis entitled

"Gap Closure to Complete the Genome Assembly for *Staphylococcus agnetis*"

written by

Joseph A. Koon II

and submitted in partial fulfillment of
the requirements for completion of
the Carl Goodson Honors Program
meets the criteria for acceptance
and has been approved by the undersigned readers.

Dr. Nathan Reyna, thesis director

Dr. Jeff Matocha, second reader

Dr. Doug Reed, third reader

Dr. Barbara Pemberton, Honors Program director

30 April 2015

Gap Closure to Complete the Genome Assembly for *Staphylococcus agnetis*

Joseph Koon, Ouachita Baptist University
Arkadelphia, Arkansas

Abstract

Poultry lameness is a significant problem resulting in millions of dollars in lost revenue annually. In commercial broilers, the most common cause of lameness is bacterial chondronecrosis with osteomyelitis (BCO). The majority of BCO infections involve *Staphylococcus agnetis*, a bacterium previously not isolated from poultry. Administering *S. agnetis* in drinking water to broilers reared on wire flooring increased the incidence of BCO three-fold when compared with broilers drinking tap water (P = 0.001). We are completing an assembly of the sequence of the *S.* agnetis genome. Currently, the assembly is in 2 large contigs of 2491 kbp and 38 kbp, and 7 small contigs of 2-3 kbp. We have been using bioinformatics and PCR sequencing to determine how these contigs are organized relative to the main bacterial chromosome and plasmids to constitute the entire *S.* agnetis genome. Annotation of that genome and identification of toxin genes will allow future researchers to better understand the etiology of BCO in broilers, and possibly develop vaccines for reducing BCO in commercial settings. Isolating pathogenic bacterial species, defining their likely route of transmission to broilers, and genomic analyses will contribute substantially to the development of measures for mitigating BCO losses in poultry.

Introduction

Every year, poultry companies lose millions of dollars because of issues associated with chicken lameness. Bacterial chondronecrosis with osteomyelitis (BCO) is the most common cause of lameness in broilers. Broilers are chickens exclusively bred for their meat. Previous research showed that the *Staphylococcus* genus was found to be the main bacterial genus involved in BCO cases (Al-Rubaye, 2013). Among the *Staphylococci*, *S. agnetis* was the most prevalent species isolated from the bone and blood samples collected from the tibiae and femora of lame broilers. It should also be noted that the aforementioned research was the first time the *S. agnetis* bacterium had been associated with poultry. Previously, only sequences for a few genes were available from *S. agnetis*, and nothing else was known about the rest of the genome. Before this research, Dr. Rhoads[1] had been working to assemble a draft sequence for the *S. agnetis* genome, along with colleagues from Oklahoma State University and the University of Arkansas. Having a completed genome sequence will put existing research done on this bacterium in context and provide ideas for future research. Current research is also focused on the epidemiology and ecology of *S. agnetis*. This includes the mode of infection (airway vs. gut), and what stresses maximize infection.

When I began working with Dr. Rhoads, the preliminary sequence of this bacterium's DNA (which was arranged in nine contigs) contained eight sequence gaps. Sequence gaps are gaps between adjacent contigs in a preliminary genome sequence. Gaps in a sequence are usually caused when those specific regions of the DNA sequence are either unstable during DNA preparation or are less amenable to the process required for Next-Generation Sequencing (Brown, 2002). Although this occurs relatively early in the Next-Generation Sequencing process, it is usually not until the end of the sequencing process that the existence of gaps is realized. The purpose of my research is to fill in these gaps in the bacterium's DNA sequence.

[1] Douglas D. Rhoads, Ph.D., is the Director of the Interdisciplinary Graduate Program in Cell and Molecular Biology at the University of Arkansas. He served as the supervising researcher throughout this project.

Experimental Procedures

Primer Design

In order to span the eight gaps in the *Staphylococcus agnetis* sequence, it was necessary to design primers that would amplify the regions of the sequence where these gaps occurred. Since we needed to amplify the sequence gaps, I designed primers that pointed away from the adjacent ends of the nine contigs. The ideal primer needed to be roughly 29 to 40 bases long, with a melting point of nearly 60° Celsius. In order to maximize the melting temperature with as few bases as possible, we also designed primers with a composition of at least 35% guanine and cytosine, since these nucleotides form triple bonds with each other, thereby giving the primer a higher melting point.

Polymerase Chain Reaction

A Polymerase Chain Reaction is a relatively complex reaction that uses several enzymes and solutions held at different temperatures to amplify specific regions of DNA. My PCR reactions were carried out in 40μL volumes. Using this volume of solution ensured that I had enough PCR product to run in a gel, purify, and sequence. For my reactions, the first step was to add 30.4μL of autoclaved millipure H_2O. I then added 4.0μL of 10X "959" Taq polymerase buffer. After adding the buffer, I added 0.4μL of clean "908" *S. agnetis* DNA at 20ng/μL, 0.4μL of 20mM dNTP solution, 0.4μL of 25mM $MgCl_2$ solution, and 0.4μL of Taq polymerase at 20units/μL. The last step was to add 2μL of both the forward and reverse primers to the PCR solution (each at 3.4μM).

Our Polymerase Chain Reactions were carried out at specific temperatures. The initial denaturation step was set at 90° Celsius for 30 seconds. The follow-up denaturation was set at 90° Celsius for 15 seconds. The primer annealing phase was programmed at 60° Celsius for 15 seconds. The elongation stage was set at 72° Celsius for 60 seconds. These last three steps were done a total of thirty times. The final elongation stage was set at 72° Celsius for 180 seconds. After the reaction was completed, the PCR machine was set to keep the samples at 4° Celsius until they could be retrieved.

Gel Electrophoresis and Typhoon FLA 9500 Operation

In order to visually see the PCR products, it was necessary to run a stained sample of the amplified DNA in a 1.0% agarose gel. To make the 1.0% agarose gel, I added 1g of electrophoresis-grade agarose per 100mL of 0.5X TBE solution. The buffer used in the gel rig was 5μL of Ethidium Bromide added with enough 0.5X TBE to cover the gel. In order to stain the DNA, I added 3μL of "Ficoll BPBXC EtBr" stain to an 8μL sample of my PCR product. I added 8μL of my stained DNA to each well. I also used 3μL of "pGEM Sau3A 24ng/μL" and 5μL of 1kb ladder as my two markers. Digested plasmid DNA was run in a 0.7% agarose gel.

After gel electrophoresis was complete, I took an image of the gel using a Typhoon FLA 9500 fluorescent imaging device. Using the fluorescence mode, the Typhoon device was set to scan for EtBr at a PMT of 550, with a pixel size of 200μm.

DNA Quantification via TKO Fluorimeter

In order to determine the concentration of DNA in my PCR product, I used a Hoefer TKO 100 Fluorometer. This fluorometer uses Hoechst dye to determine the change in concentration of the solution when a DNA sample is added it. Before using the fluorometer, I always turned the device on and let it warm up for thirty minutes. During this time, I made the Hoechst dye solution. The solution is 1μL of Hoechst dye per 10mL of 0.5X TBE buffer. After the fluorometer had time to warm up, I added 2mL of the Hoechst solution into a cuvette and placed it in the device. After waiting a few seconds, I turned the "zero" knob until the display read "0 units". I then added 2μL of *E. coli.* DNA with a standard concentration of 60ng/μL to the cuvette, mixed the cuvette very well, and turned the "scale" knob until it read "60 units". At this point, the fluorometer was calibrated and could be used to determine the concentration of DNA in my samples. I then added 2μL of my sample to the cuvette, mixed the cuvette, and subtracted the display reading by sixty units to get the concentration of DNA in my sample in ng/μL. For example, if after adding 2μL of my sample to the cuvette, the display read "86 units", then the concentration of DNA in my sample was 26ng/μL.

Purification of PCR Product

Before sending my PCR product to get sequenced, it was necessary to purify the product. We found that an easy and effective method for purification is to use the *RapidTip* ® method. This purification method removes dNTPs, primers, and primer-dimers that are left over after PCR. In order to purify a sample, I set the pipette to 80μL and, using a *RapidTip* ®, I aspirated the sample into the tip and then pushed the entire sample back into the same tube. This was repeated two more times and then I moved to a fresh tube and continued the step two more times, meaning that the sample passed up and down through the tip a total of five times.

Preparing PCR Products for Sanger Sequencing

After purifying my PCR product, each sample was equally divided into two tubes. I then added one of the two primers used in the PCR reaction to each tube. By using both primers in the Sanger sequencing process, I was able to get two new DNA sequences from one DNA fragment. The total volume of the PCR product, primer, and millipure H_2O was always 13μL.

Sanger Sequencing

Sanger Sequencing for this project was done in the Poultry Science building on the University of Arkansas campus.

(Quick) Assembly

After receiving new sequences, the first step was to attempt an assembly of the new sequences with the existing contigs of *S. agnetis*. This was done using SeqMan Pro software. If any of the sequences did not readily fit into any of the existing contigs, the next step was for me to clean up the sequence and try again.

Clean-Up of MiSeq Data for Scaffold Construction

Sometimes, the computer software used in Sanger sequencing fails to clearly read the peaks made by the fluorescently labeled ddNTP's. These errors can sometimes keep new DNA sequences from being assembled into the existing contigs. Cleaning up the MiSeq data is done using EditSeq software, and has two parts. The first part is to scan through the sequence to see if any of the peaks were improperly labeled. The second step is to try extending the ends of the sequence to incorporate more bases than the sequencing software might deem clearly discernable. By doing these two steps, new sequences will usually assemble into existing contigs, sometimes spanning sequence gaps in the process.

Bacterial Culturing

Before we could isolate plasmid DNA from the S. agnetis cells, it was first necessary to culture the cells. The bacterial cells were placed in Tryptic Soy broth and incubated at 37° Celsius until the cells were replicating in the log phase. By culturing these cells until they reached the log phase of replication, we insured that the plasmid DNA we retrieved from the cells had a high enough concentration for the following restriction step.

Plasmid DNA Isolation (includes SDS and SDS + ReadyLyse methods)

In order to lyse the *S. agnetis* cell while also preserving the plasmid DNA contained inside, Dr. Rhoads used the below procedures.

Alkaline SDS

Before beginning this isolation protocol, the *S. agnetis* bacteria was incubated overnight at 37°C in tryptic soy broth (TSB), allowing the bacteria to reach the log phase of replication. After the overnight incubation, the first step was to dilute the culture 100-fold with TSB and incubate at 37°C for 3 hours in a shaker. The solution was then separated into 25mL tubes, and pellet the cells @ 8,000rpm for 5 minutes. The pellet was then resuspended in 1mL of TE, and transferred to a microcentrifuge tube. The solutions were then pelleted @ 8,000 rpm for 5 minutes. The pellets were then resuspended in 100μL of TNE. At this point, 350μL of 150mM NaOH 1% SDS was mixed into the tube and incubated for 5 minutes at 37°C. Then 260μL of 5M KoAC was added and mixed into the solution. The solution was then spun for 2 minutes at 10,000rpm at 4°C. The supernatant was transferred to a new tube, and DNA extraction was done with 50:50 phenol $CHCl_3$. The resulting top aqueous solution was then decanted and mixed with an equal volume of isopropanol. The solution was then spun for 5 minutes at 11,000rpm at 4°C. The solution was then decanted, rinsed with 70% EtOH, spun for 1 minute, decanted, and left to dry. The pellet was redissolved in 90μL of TE buffer, with RNAse A (at 20μg/mL), and incubated at 37°C for 30 minutes. Another extraction was then done with $CHCl_3$. To this solution, 10μL of 3M NaOAc and 250μL of cold EtOH were added. This solution was then spun for 5 minutes at 11,000 rpm at room temperature and decanted. After this, 200μL of cold 70% EtOH was added to the solution. The solution was then spun, decanted, dried, and resuspended in 30μL of Te. The resulting isolated plasmid DNA was then quantified.

Alkaline SDS + ReadyLyse

Before beginning this isolation protocol, the *S. agnetis* bacteria was incubated overnight at 37°C in tryptic soy broth (TSB), allowing the bacteria to reach the log phase of replication. After the overnight incubation, the first step was to dilute the culture 100-fold with TSB and incubate at 37°C for 3 hours in a shaker. Then separate the solution into 25mL tubes, and pellet the cells @ 8,000rpm for 5 minutes. The pellet was then resuspended in 1mL of TE, and transferred to a microcentrifuge tube. The solutions were then pelleted @ 8,000 rpm for 5 minutes. The pellets were then resuspended in 100µL of TNE. After the resuspension, 2µL of Ready Lyse was added to the solution, and it was incubated at room temperature for 20 minutes. At this point, 350µL of 150mM NaOH 1% SDS was mixed into the tube and incubated for 5 minutes at 37°C. Then 260µL of 5M KoAC was added and mixed into the solution. The solution was then spun for 2 minutes at 10,000rpm at 4°C. The supernatant was transferred to a new tube, and DNA extraction was done with 50:50 phenol $CHCl_3$. The resulting top aqueous solution was then decanted and mixed with an equal volume of isopropanol. The solution was then spun for 5 minutes at 11,000rpm at 4°C. The solution was then decanted, rinsed with 70% EtOH, spun for 1 minute, decanted, and left to dry. The pellet was redissolved in 90µL of TE buffer, with RNAse A (at 20µg/mL), and incubated at 37°C for 30 minutes. Another extraction was then done with $CHCl_3$. To this solution, 10µL of 3M NaOAc and 250µL of cold EtOH were added. This solution was then spun for 5 minutes at 11,000 rpm at room temperature and decanted. After this, 200µL of cold 70% EtOH was added to the solution. The solution was then spun, decanted, dried, and resuspended in 30µL of Te. The resulting isolated plasmid DNA was then quantified.

Restriction Enzyme Digestion

In order to do the restriction enzyme digestion, the following protocol was used. When using the HindIII restriction enzyme, the first step was to add 2µL of 10X React 2 buffer to the tube. Then 2µL of BSA at a concentration of 160mg/mL was added to the tube. The next step was to add 12.5µL of millipure H_2O to the solution. Next, 3µL of the DNA sample were added to the tube. Using a concentration of 10units/µL, only 0.5µL of restriction enzyme were added to the solution. When using the EcoRI restriction enzyme, the last step was to add 2µL of 10X React 3 buffer.

Results and Discussion

In order to determine the unknown DNA sequence in the gaps, I have had to utilize a variety of protocols. After becoming familiar with the previous work Dr. Rhoads has done on *S. agnetis*, I designed a variety of primers that point away from the adjacent ends of these nine contigs. After we received these new primers in the mail, Dr. Rhoads determined the best protocol to test all possible primer combinations. After I did a PCR on each of the 241 primer combinations and ran a gel of the amplified regions, Dr. Rhoads determined the 22 most promising combinations for sequencing (see FIG. 1. and 2.). After sequencing, we had 44 new DNA sequences with which to consolidate contigs and build scaffolds. When designing these new primers, I had hoped they would produce reads that would span all of the eight gaps in the *S. agnetis* sequence. However, Dr. Rhoads did a quick assembly of the 44 new sequences and showed that the primers I designed were not successful in spanning all of the eight gaps.

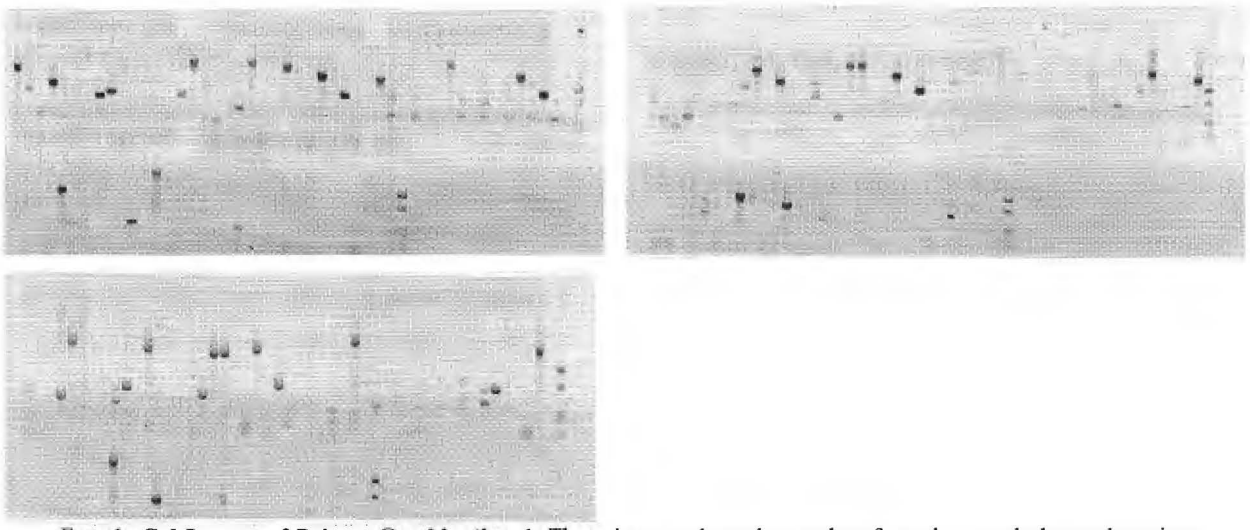

FIG. 1. **Gel Images of Primer Combination 1.** These images show the results of running a gel electrophoresis on each of the 241 primer combinations. If the two primers in each sample annealed at locations on the *Staphylococcus agnetis* sequence that were conducive to PCR amplification, then a band would be visible in these gel images. From these 241 samples, Dr. Rhoads and I chose the 22 primer combinations that successfully amplified a region of the *S. agnetis* genome.

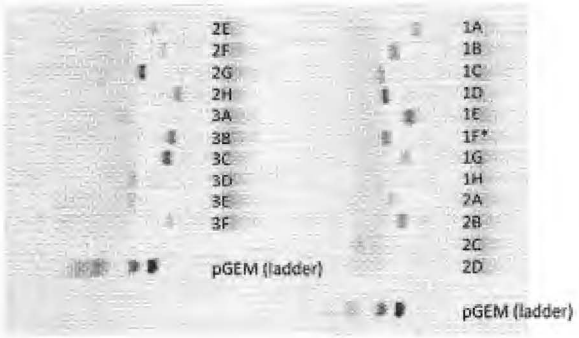

FIG. 2. **Gel Image of the 22 Effective Primer Combinations.** This image shows that the 22 primers we selected were indeed effective at annealing to and amplifying a region of the *S. agnetis* genome. Each of these 22 PCR product samples was purified using the *RapidTip* ® method, divided equally into 44 wells, combined with one primer, and sent off for sequencing. This produced a total of 44 new DNA sequences which were used to construct new scaffolds, combine contigs, and close sequence gaps.

The new sequences we got from applying these primers to the *S. agnetis* DNA also yielded five large repeated sections (see FIG. 3.). Dr. Rhoads determined that these repeated sections were the 16S and 23S genes. Since it was previously shown that multiple copies of the 16S gene could exist in a genome, these results were not abnormal (Case, Boucher, Dahllof, Holmstrom, Doolittle, & Kjelleberg, 2007).

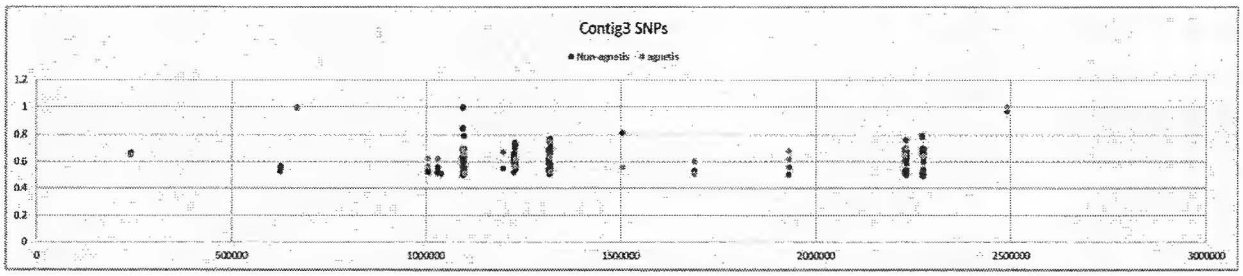

FIG. 3. **Single Nucleotide Polymorphisms in *S. agnetis* and similar *Staphylococcal* species.** Each dot on this graph represents a SNP (single nucleotide polymorphism). The black dots are the locations of SNPs in *Staph.* species similar to *S. agnetis*. The gray dots are the location of SNPs in *Staph. agnetis*. The x-axis is in number of nucleotides, which tells us what position each SNP occurs in the sequences. The y-axis is the percent of SNP in the two samples. When both dots appear together, it can be inferred that these are regions where SNPs usually occur in these *Staph.* species. Dr. Rhoads determined that the five large conserved SNP regions are 16S and 23S genes.

I then discovered that a 38kbp contig contained two copies of a large repeat (>10kbp). In order to investigate the possibility that the last 10kbp of this large contig was actually the same as the first 10kbp, Dr. Rhoads and I purified the plasmid using three different methods. The first method (method A) combined 150mM NaOH 1% SDS with a Ready Lyse enzyme. The second method (method B) only used 150mM NaOH 1% SDS. The isolated plasmids were then digested with restriction enzymes. We then compared the gel image from these digestions to see if the enzymes cut where the sequence predicted when viewing the sequence as a plasmid. The gel image showed that the plasmid purification step was not clean enough to obtain clear results (see FIG. 4.). In order to acquire a clearer gel image, Dr. Rhoads used the Lysostaphin lytic enzyme with the Alkaline SDS protocol from the first two methods (see FIG. 5.).

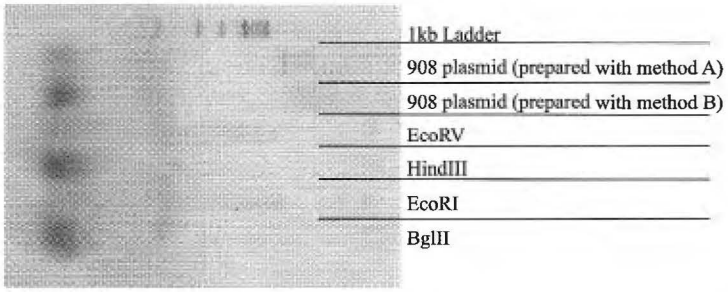

FIG. 4. **Gel Image of Digestion 1.** This image shows that the first two plasmid purification protocols we used did not yield a clear gel image. "Method A" used 150mM NaOH 1% SDS combined with a Ready Lyse enzyme. "Method B" only used 150mM NaOH 1% SDS.

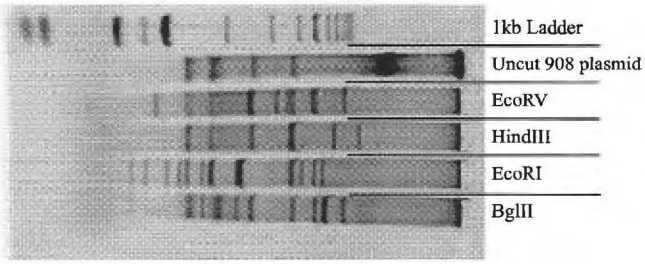

FIG. 5. **Gel Image of Digestion 2.** This image shows that using the new Lysostaphin lytic enzyme in combination with NaOH 1% SDS is a successful method for isolating plasmid DNA from *S. agnetis* cells.

At this point, I continued to design a variety of primers to PCR amplify regions of DNA in the remaining gaps between contigs in the *S. agnetis* draft genome. Since there are several repeated sections in the *S. agnetis* genome, I had to make sure the new primer sequences were specific for the regions that I wanted to amplify. FIG. 6 shows that from nine new primer combinations, only one of the combinations successfully amplified a region of DNA.

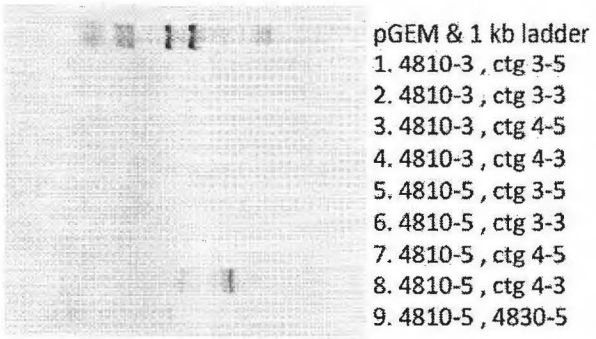

FIG. 6. **Gel Image of Primer Combination 2.** In the first lane, I ran the pGEM and 1kb ladders together. The results of this gel were surprising because I had expected most of these primer combinations to amplify *S. agnetis* DNA. After seeing this gel, I repeated the same PCR reactions in order to confirm these results. I used the primer combination labeled number 8 above as the positive control for the second reaction.

Since the results from these PCR amplifications were unexpected, I decided to repeat this step with the same primer combinations to ensure that I did not make any errors in setting up the PCR reaction.

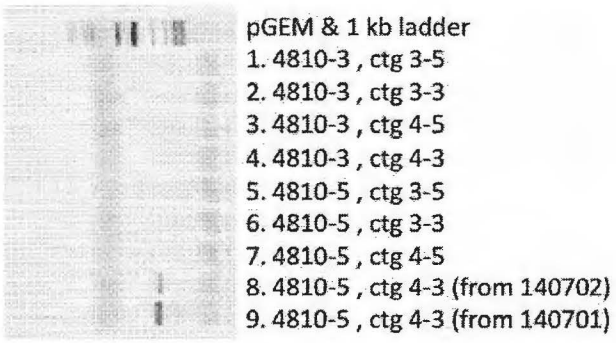

FIG. 7. **Gel Image of Primer Combination 2 (redo).** In the first lane, I ran the pGEM and 1kb ladders together. As an extra precaution, I repeated the reaction from combination 8 in FIG. 6. above and ran both the new and old PCR products as reactions 8 and 9, respectively. Since the same results were seen in combinations 8 and 9, I could confirm that primer combinations 1 through 7 did not amplify any regions of DNA in the *S. agnetis* genome.

After confirming that only one of these primer combinations was able to amplify any *S. agnetis* DNA, I proceeded to purify the PCR product from reactions 8 and 9 in FIG. 7. When I had purified and quantified the DNA, I sent the PCR product off for Sanger sequencing in both directions and incorporated the two new sequences into the *S. agnetis* genome. By adding these new sequences to the draft genome, I was able to combine two contigs together.

Since the incorporation of these two sequences into the genome was not sufficient to close all of the gaps, Dr. Rhoads used the DNA Navigator program to design four more primers. All of these new primers annealed to and amplified regions of the *S. agnetis* sequence (see FIG.

8). Since all four primer combinations produced a PCR product, their products were purified, quantified, and submitted for Sanger sequencing in both directions.

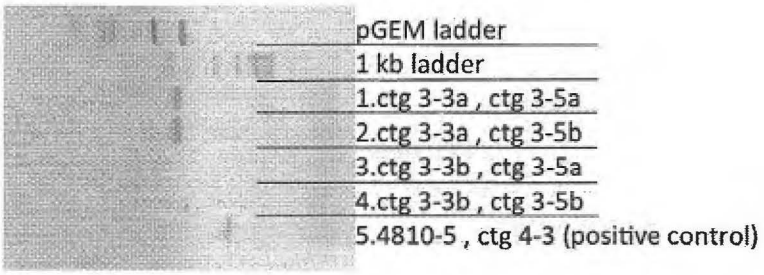

FIG. 8. **Gel Image of Primer Combination 3.** In the first lane, I ran the pGEM ladder, and the 1kb ladder was run in the second lane. The same positive control used in FIG. 6 and 7 was used here as well. Since all of these primer combinations amplified regions of the *S. agnetis* genome, they were all submitted for sequencing.

After I received my sequences, I noticed that the MiSeq data showed very few, if any, clear peaks. In an attempt to resolve the issue, I repeated these PCR reactions and resubmitted the products for sequencing. Unfortunately, these new sequences also had very few clearly discernable peaks. I talked with Dr. Rhoads about this issue, and he informed me that these results were characteristic of having a primer that was self-amplifying in PCR. In other words, ctg 3-3a and ctg 3-3b were able to amplify DNA on their own (i.e. without working in conjunction with another primer). This explains why reactions 1 and 2, and 3 and 4 produced the same sized band. At this point, my research term was over, so I was not able to correct any of these primers.

Conclusion

At the beginning of this project, the *Staphylococcus agnetis* genome was in nine contigs. As a result of these efforts, the genome is now in four contigs. In the future, this lab hopes to design more primers that will amplify regions of DNA that will be used to fill in the remaining sequence gaps.

Acknowledgments

This project was supported by the Arkansas INBRE program, with grants from the National Center for Research Resources - NCRR (P20RR016460) and the National Institute of General Medical Sciences - NIGMS (P20 GM103429) from the National Institutes of Health.

References

Al-Rubaye, A. A. (2013). Molecular Diagnosis of Metabolic Fast Growth Related Diseases in Broiler. *ProQuest* .

Brown, T. A. (2002). Sequecing Genomes. In *Genomes* (2nd Edition ed.). Manchester, United Kingdom: Oxford: Wiley-Liss.

Case, R. J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W. F., & Kjelleberg, S. (2007). Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies. *Applied and Environmental Microbiology , 73*, 278-288.

Smeltzer, M. S., & Gillaspy, A. F. (2000). Molecular Pathogenesis of Staphylococcal Osteomyelitis. *Poultry Science* , 1042-1049.

Wideman Jr., R. F., Al-Rubaye, A., Gilley, A., Reynolds, D., Lester, H., Yoho, D., et al. (2013). Susceptibility of 4 Commercial Broiler Crosses to Lameness Attributable to Bacterial Chondronecrosis with Osteomyelitis . *Poultry Science* , 2311-2325.

Wideman Jr., R. F., & Prisby, R. D. (2013). Bone Circulatory Disturbances in the Development of Spontaneous Bacterial Chondronecrosis with Osteomyelitis: A Translational Model for the Pathogenesis of Femoral Head Necrosis. (M. S. Cooper, Ed.) *Frontiers in Endocrinology , 3*, 1-14.