Closing the Gap:
Determining the Nature of a Gap in Arhodomonas sp. Seminole DNA to Predict Gene Coding

GRP# 16
Section# 002

Mark Castaneda, Micah Gillezeau, Madison Slawson, Matthew Zembraski, Dr. Patricia Canaan, Ph.D.

ABSTRACT

The purpose of our research is to complete the genomic sequence of Arhodomonas sp. Seminole so that we can identify the gene(s) that allow this species to survive in its environment and metabolize the crude oil. Arhodomonas sp. Seminole has 750 pieces that comprise its genomic DNA. We want to identify how they link together. Our group was given two of those pieces and charged with determining how they related to each other. Our hypothesis is that the contigs link together and code for a gene. In order to test our hypothesis we performed several tests including analysis on Blast, a database of the National Library of Medicine, PCR amplification, gel electrophoresis, and sequencing DNA. From our data we discerned that our two contigs did overlap one another and that the joined contigs likely coded for a known protein, 2-keto-gluconate dehydrogenase. From our data, we were able to conclude that our hypothesis is supported because our amplified DNA produced a product in the gel electrophoresis, and the sequenced DNA aligned with many known genes identified in other organisms.

INTRODUCTION

Arhodomonas sp. Seminole was discovered at an oil-site in Seminole County, OK. The species is unique in the fact that it can survive high levels of NaCl and is able to grow by consuming many different carbon sources (Canaan, 2014). Arhodomonas sp. Seminole is able to degrade benzene and toluene among other compounds as the sole carbon source at high salinity (Canaan, 2014). As a result of these unique characteristics, Arhodomonas sp. Seminole is a fitting organism to use at crude oil drilling sites. Our purpose is to obtain a completed genomic sequence for Arhodomonas sp. Seminole. We are working to finish the genomic sequence of Arhodomonas sp. Seminole by closing gaps between contigs. The genomic sequence has 750 pieces, or contigs. Determining how the contigs are joined together will help us complete the genomic sequence and determine which genes allow this species to survive in its extreme conditions. Our ultimate goal for the research is to determine what makes Arhodomonas sp. Seminole capable of metabolizing the crude oil. There are no other known species that are able to metabolize the crude oil at drilling sites in the presence of such high concentration of NaCl. Therefore, Arhodomonas sp. Seminole is of high interest to scientists who wish to preserve sustainability of our land. We hypothesize that the contigs we were provided join and code for a gene.

MATERIALS AND METHODS

To start the experiment, we received two contigs of DNA. We used BLAST, an online database of known proteins, to compare our DNA to the known protein *tr|A1KCD3|A1KCD3_AZOSB Putative glucase dehydrogenase alpha subunit OS=Azoarcus sp. (strain BH72) GN=azo3873 PE=4 SV=1* to predict whether the contigs joined. We first compared our contigs as separate pieces and then a fused contig with N's between the contigs to represent the predicted gap. We then began the process of PCR amplification in order to replicate and amplify the gene of interest. To create PCR primers, we first generated a subsequence comprised of 200 bases upstream and downstream of the gap in our fused contig. We then used IDT to design two primers, a forward and reverse, and predict the size of a product. Once we received the manufactured primers, we created a PCR solution in order to amplify our DNA. The 100 µl PCR mix contained 50 ng of DNA,0.6 µM of each primer, 200 µM of PCR nucleotide mix (Fisher Bioreagents, Fisher Scientific, Pittsburgh, PA),1.75 mM MgCl2, 2.5 units Taq polymerase in Buffer A (M1865, Promega Chemicals, Madison, WI). The amplification was performed in a thermocycler with a program which consisted of an initial denaturation at 94˚ C for 30 seconds, 30 cycles of denaturation at 94˚ C for 30 seconds, annealing at 55˚ C for 30 seconds, extension at 72˚ C for 1 minute, and a final 10 minutes at 72˚ C for the final extension. We took a 10 µl sample of the amplified DNA, added a tracking dye, and then ran an agarose gel electrophoresis. We sequenced the DNA of our primers, and then used Clustal to compare the primers to our fused contig to identify the nature of the gap (European Bioinformatics Institute). Finally, we used BLAST to identify known proteins that aligned with our fused contig to predict what the gene codes for.

RESULTS

The first blast showed us that the two contigs likely join because the protein aligned with both contigs without any breaks in the amino acids. When compared to the fused contig, no amino acids aligned with the gap of n's, and the protein still aligned with the contig. Having sufficient evidence to proceed, we moved to designing PCR primers to facilitate PCR amplification. We selected the forward primer CATGCAGGAGCACGAGTT and reverse primer CCGTCATAGGGCTTGGTATTG, which IDT predicted would yield an amplicon length of 279. When we performed the gel electrophoresis (**Figure 1**), the product was less than 300 base pairs long (**Figure 2**).

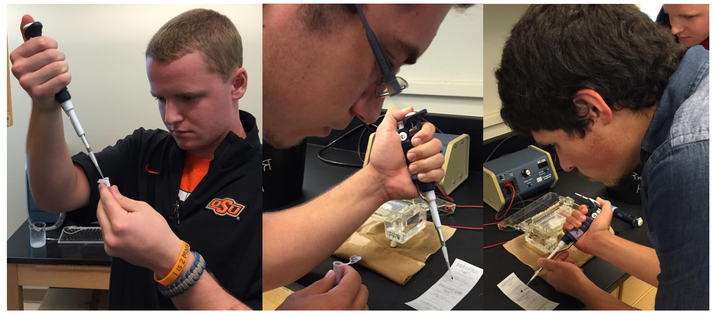


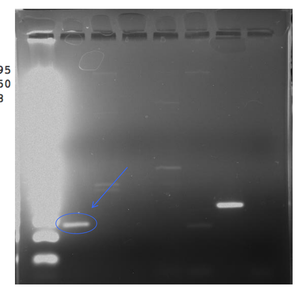
Figure 1: From left to right, Matthew extracts the PCR solution, Mark places it on the wax paper, and Micah adds the tracking dye to the solution. The solution was then placed in the agarose gel.

When we compared our sequenced primers to the fused contig, we saw that both the forward and reverse primers aligned to the fused contig on both sides of the gap (**Figure 3**). We also see that each primer has an adenine in the gap, and the fused contig has the sequence adenine-guanine-adenine downstream of the gap that does not align with either of the primers (**Figure 3**).

```
16R_rc                    CCGAGA-----------AGAAGATGGGGGTGACCGGCAGCAAGGAGACC 195
Contig671_tail_Contig465_head  CCGAGNNNNNNNNNNAGAAGATGGGGGTGACCGGCAGCAAGGAGACC 550
16F                       CCGAGA-----------AGAAGTTGGGGTTGACCGGCAGCTTGGAGACC 68
                          *****           ***** ***** ***********  *******
```

Figure 3: Isolated region of the alignment between the primers and the fused contig, centered around the gap of n's. *Source: European Bioinformatics Institute.*

Figure 2: The agarose gel displays our PCR product between the second and third band of the control.



Our experiment was limited in that our process only allowed for a single trial to be performed. To ensure that our results were not the result of luck, we should perform three to four more trials of the PCR amplification, gel electrophoresis, and DNA sequencing. However, having an experienced team, comprised of our professor and graduate teaching assistants, guide us through the process allowed us to ensure that our experiment was largely without error. We were able to meet the objective of our experiment, joining two of the contigs within the genome. Canaan and her colleagues were able to predict many of the genes present in the Arhodomonas genome. Through our experiment and the continuation of it with other contigs of the genome, researchers will be able to confirm the presence of these genes in the genome.

DISCUSSION

Through the first blast, we were able to provide evidence that the two contigs are connected as we hypothesized since the given protein aligned without breaks. Because none of the protein's amino acids fell within the gap of the fused contig, we concluded that no gap existed between the contigs, suggesting that the contigs either overlap or are blunt. Because we had reasonable evidence to suggest that the contigs join, we designed upstream and downstream primers to allow us to amplify the gene of interest within the DNA and have enough copies of said gene for gel electrophoresis and sequencing. The band on the gel was just less than 300 nucleotides long, which suggests that our primers worked since the predicted amplified region length was 279. Our final blast of the forward and reverse primers with the fused contig showed that the two contigs overlap because two nucleotides downstream of the gap do not align with either primer. Our final objective was to determine the purpose of this sequence of DNA. We determined that our portion of the genomic sequence likely codes for the protein 2-keto-gluconate dehydrogenase, which is found in Pseudomonas stutzeri among other organisms. According to MetaCyc, 2-keto-gluconate dehydrogenase catalyzes the metabolism of 2-dehydro-D-gluconate into D-gluconate using NADPH as an energy source (**Figure 4**). This is part of the ketogluconate metabolism pathway (**Figure 5**). Going forward, we would like to continue towards our goal of creating one contiguous genomic sequence and isolate the gene(s) that allows Arhodomonas sp. Seminole to survive and the thrive in its extreme environment with hopes of finding and/or creating other organisms that are able to metabolize crude oil.

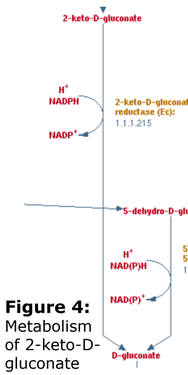


Figure 4: Metabolism of 2-keto-D-gluconate

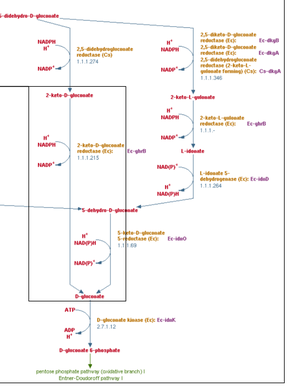


Figure 5: Ketogluconate metabolism pathway

REFERENCES

- *BLAST.* National Center for Bioctechnology Information. Web. 3 Sept. 2014.
- Canaan, Dalvi, Nicholson, Najar, Roe, Hartson, Fathepure. "Arhodomonas sp. Strain Seminole and Its Genetic Potential To Degrade Aromatic Compounds under High-Salinity Conditions." *Appl Environ Microbiol.* 2014 Nov 1: 80(21):6664-76. *American Society for Microbiology* 2014 Aug. 22. Web. 10 Nov. 2014.
- *European Bioinformatics Institute.* European Molecular Biology Laboratory. Web. 14 Nov. 2014.
- *IDT Integrated DNA Technologies*. Integrated DNA Technologies, Inc. Web. 17 Sept. 2014.
- "MetaCyc Pathway: ketogluconate metabolism." *MetaCyc.* BioCyc. Web. 31 Oct. 2014.