Predicting the DNA Sequence of *Arhodomonas sp. seminole*

GRP #21
Section#702

Canaan, Patricia; Hearon, Kyleigh; Markovics, Emally; Perry, Stormy; Peterson, Whitney

ABSTRACT

In this experiment we dealt with the bacteria *Arhodomonas sp. Seminole*, which is an aerobic, halophilic bacterium. Our goal is to sequence the DNA so we can identify and predict the enzymes, traits, phenotype, and other aspects of the bacterium. The DNA was extracted from the bacterium by rupturing cells, purifying the DNA, and then concentrating it by precipitation. DNA is found in the genes of all living organisms. When there would normally be one whole contiguous piece of DNA (complete genome), our genes are separated into two pieces (called contigs). There are missing segments within the gene and our job is to determine what is missing and how we can fit it together to form a complete genome. We obtained a specific region where we generated DNA sequence data of a gap to bring two pieces together. We also aligned our related protein to our fused contigs to determine if they could possibly be joined. We used procedures such as Polymerase Chain Reaction (PCR), thermocylcling, and DNA Gel Electrophoresis to obtain our results. We also used online methods such as BlastX and ClustalW2 that sequences our contigs. Our results proved the hypothesis that our contigs were joined. We obtained a reverse sequence and when we related it to the fused contig, there were matches before the gap as if it was a forward primer, proving that the contigs were joined.

INTRODUCTION

Arhodomonas sp. Seminole is an aerobic, halophilic bacterium that was enriched from salty, crude-oil-impacted soil in Seminole Co., Oklahoma. It grows on a wide variety of carbon sources and requires salt. It has been shown that it can metabolize benzene, toluene, phenol, 4-hydroxybenzoic acid (4-HBA), protocatechuic acid (PCA), 12 and phenylacetic acid (PAA) as the sole sources of carbon at high salinity (1). We started with fused contigs and a related protein and our goal was to determine if the upstream and downstream contigs aligned with the protein and if they were joined. Our hypothesis was that the two contigs were aligned with some unknown protein between them. Our results proved the hypothesis and our contigs were in fact adjacent. Our overall purpose was to complete the sequencing for the genome for this particular bacterium so that we may learn more about its molecular characteristics.

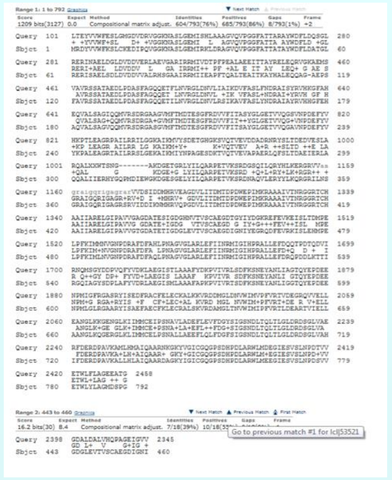
MATERIALS AND METHODS

Small glass tube used for rupturing of cells; Glass beads; Arhodomonas DNA fragments; The sequence of a full length contig, 21 related protein sequences, and 21 fused contigs; Sequence for a related protein; P20 Micropipette; Test tube, Disposable tips; Tag polymerase composed of the following: dH2O (provided 70 µL),10x Taq buffer (10 µL), 10x dNTP's (all 4 10 µL), Arh. Sp. Seminole (5 µL), gDNA (58ng/µL), F-primer (2 µL), R-primer (2 µL), Taq Polymerase (1 µL); Agarose Gel electrophoresis; Reagents; Arhodomonas DNA; Proteins; Thermocycler; 8 µm of PCR sample; 2 µL of blue, tracking dye.

The bacterium used in this project is Arhodomonas sp. Seminole, which can be found in salty, crude-oil impacted soil in Seminole, OK. On September 3, 2014 we were given a sequence gap to examine and identify, this sequence gap came from the DNA that was extracted from the Arhodomonas. We then obtained word documents containing The 21 full length contigs, 21 related protein sequences, and 21 fused contigs. We took these given documents and used BLASTX to find the alignments in the DNA between one contig and the related protein. Once we found our genes and regions of interest in our given contigs, on September 10, 2014 we designed a PCR amplification of the sequence gap using BLASTx. Once the PCR amplification was designed, using a micropipette, we set up the PCR to amplify the gap regions by transferring 70 ul of dH20, 10 ul of 10x tag buffer and 10x dNTP's, 5ul of arhodomonas sp. Seminole gDNA, 2 ul of F-primer and R-primer and 1 ul of Taq polymerase into a test tube. It was then sent for DNA sequencing using the thermocycler. The thermocycler sent the DNA through a variation of temperatures , which resulted in copies of fragments. Once we retrieved our copies of fragments we transferred them into the agarose gel electrophoresis using a micropipette. Using the results from the electrophoresis, we examined the genomic DNA sequence around (flanking) the gap region and then finalized our DNA sequence analysis. After getting our results which showed us that our hypothesis was supported, we used blast x to blast our fused contigs to find a specific protein that fit our contigs. In our contig we had a "good match" and found that phosphoenolpyruvate synthase was our protein.

RESULTS

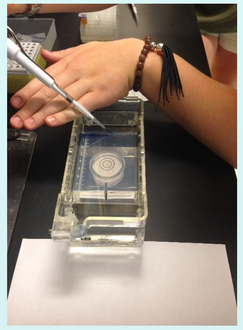
We were given two DNA sequences from the Bacterium Arhodomonas Sp. Seminole and we hypothesized that these two contigs were related sequences and connected through a gap. Our first experiment were placed these two contigs through the BLASTX Program to determine whether these two sequences aligned. Our Results were as followed.

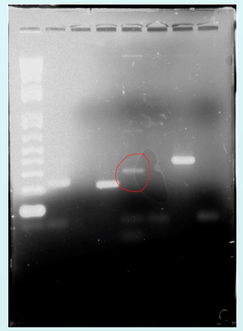


Thanks to the Blastx program we saw how our two contigs patterns had the potential to align well enough for us to create a reverse primer to complete the pattern. We accumulated a few cells and ruptured them in order to purify the DNA using a concentrate by precipitate.

Our next procedure involved the taq polymerase. We mixed our purified DNA sequence of Arhodomonas Seminole, and added our Forward Primer, Reverse Primer, our taq buffer, H2O, and our taq polymerase. Using a micropipeter we placed all these items in a single tube, and put them through a centrifuge for approximately four hours.

After placing our PCR product in an Agarose Gel Electrophoresis we found that our reverse primer had made a product, but our forward primer did not. The following is the picture of our gel products. Our gel is the fifth from the left.





After length of discussion as to why our product was a weak one we ran one more sequence through the Clustal program for our primers in order for us to see if we had succeeded in our prediction of the primers needed.

.

This showed us that out two contigs were indeed adjacent to each other, even though they didn't line up completely.

.

DISCUSSION

Throughout this experiment, our group predicted that our two contigs of the Arhodomonas aquaeoli were in fact adjacent to each other with a gap sequence in between. Our hypothesis was supported when we used the BLASTx method to see if the tail of the first contig and the head of the second contig aligned with a related protein, which in fact did result in an alignment between the two contigs and the related protein. The next step of supporting our hypothesis came with the PCR procedure. After priming a complete strand of DNA of the Arhodomonas aquaeoli to target a specific area that theoretically contained our two contigs, we underwent PCR to make multiple copies of this target area, which hopefully contained the sequencing needed to fill our gap between the hypothesized adjacent contigs. With a successful PCR product, we discovered that the product contained the sequencing annealed by the reverse primer, which gives us proof that the two contigs are in fact adjacent to each other in the full fused DNA strand. Although we did not get quality gene sequencing of the gap, the reverse primer gave us sequencing of the strand downstream of the gap towards the tail of the first contig. Because of this, we can conclude that the two strands are in fact adjacent to each other. If they were not, the reverse primer would only be able to anneal the sequence to the head of the second contig (upstream of the gap) and could not continue past the gap towards the tail of the first contig. Therefore, with the success of our PCR, we obtained strong, supporting evidence that the two contigs are indeed fused together despite the unknown gap in between. In addition, after running the related gene protein through BLAST, we found a match to a reference protein that contains similar sequences in multiple organisms, known as phosphoenolpyruvate synthase, which is also referred to as pyruvate, water dikinase. This protein functions as an enzyme in the process of metabolizing pyruvate, specifically in CO2 fixation. This data suggests that the two contigs contain gene sequencing for this specfic pyruvate enzyme function.

REFERENCES

• Dalvi S, Nicholson C, Najar F, Roe BA, Canaan P, Hartson SD, Fathepure BZ, 2014. *Arhodomonas sp. Strain Seminole and Its Genetic Potential To Degrade Aromatic Compounds under High-Salinity Conditions*. Applied and Environmental Microbiology. 80:6664-76